

Press tells it nearly like it is

Can the United States survive a financial crisis and an election in the same year, and still keep its science enterprise intact?

THE science enterprise in the United States will this year pay a high price for the US government's profligacy over the past eight years. So much is clear from the noises emerging from the Congress. The administration's brave plan to double the budget of the National Science Foundation over a period of five years will be further postponed. The trouble now is the pact between the administration and the Congress that will limit discretionary spending in the financial year beginning in October to just over \$3,000 million, less than necessary to match inflation. (Defence spending is separate.) The administration wants most of such extra spending as there will be to be spent on science and technology. Congress, with an election in the offing, is going to push for more popular causes instead. This could be the first year since 1980 when spending on basic science marks time.

That is the sombre background of Dr Frank Press's address to the US National Academy of Sciences (of which he is president) last week. His tale may have been more cheerful than that of his British counterpart, Sir George Porter, a few weeks earlier, but even well-endowed enterprises may be seriously hurt when growth is brought suddenly to a halt. Press, last week, was chiefly concerned to anticipate the damage that may be done, and to fend it off. His case is the more persuasive because he acknowledged that US basic science is now in good shape because the government has invested generously in its support since almost the beginning of the Reagan presidency. The issue is that of how best to limit the damage that might be done.

Press also last week accepted that the federal budget deficit must begin to shrink. The uneasy budgetary agreement between the administration and the Congress is meant to accomplish just that. Press is right to say that the US government has no choice. There are some who will say that Press is wrong to acknowledge, as he did last week, that the Congress will not fall in with the administration's wish that virtually all of the \$3,000 million to spare in the agreed budget should be spent in the cause of basic science, but that is too puritanical a view. Congressmen (as distinct from Senators) must stand for re-election every two years precisely because of the calculation by those who framed the US constitution that they would thus be responsive to their electors' wishes. It is unrealistic to complain when, as on this occasion, congressmen decide that their constituents would prefer not to see the bulk of the funds available spent on science. It might well be different if congressmen were elected for longer terms of office, but there is not much time to legislate for that between now and November.

How then to limit the damage? Press asks that the scientific community should take the initiative in helping politicians to decide where to spend the funds available. That must surely be the correct strategy to follow. It may well be that many of the problems now afflicting British science would have been less daunting if the community of researchers had been more willing to acknowledge that their paymasters also had a problem or two to solve. Press's own recipe for damage-limitation has the particular virtue of being virtuously altruistic:

he asks that, in scarcity, there should be "absolute priority" for spreading the funds available on training grants, fellowships and research grants reaching as many researchers as possible. Throw in national crises (AIDS) and notable opportunities (superconductivity, for example) and the Congress should know where to put the funds at its disposal.

The grant-making agencies would naturally find the going a little rougher. They will (or should) applaud Press's notion that the "pipeline" that supplies the talent for the research enterprise should remain full, but they may find that people appointed to research fellowships without a little money in their laboratory administrators' pockets to keep them out of mischief will eat out their hearts in discontent. But what matters is that the goal is laudable, and that its modesty should win support in the Congress even in an election year.

What comes next? People making public speeches listing categories of problems to be dealt with usually take care that there should be at least three of them. Not so Press, who lumps the rest of federally sponsored research and development into a single category of projects to be authorized now, but on which real money will not be spent until the US government has dealt with its financial problems. Building the Superconducting Super-Collider is one of these. Sequencing the human genome is another. The space station rates a mention only as an ingredient of the subcategory called the "political category". "It may be wise", Press said last week, "for multibillion-dollar projects like (*sic*) the space station to be left for major funding decisions by the next president, who . . . will have the responsibility for seeing them done." In Washington's argot, that is a way of saying that the space station is, at present, at best a waste of money and probably a diversionary exercise (and thus actively harmful) as well.

Press's enemies will probably join his friends in asking whether the people who matter most will listen to him. His enemies will say that it is not for the likes of him to tell those elected to public office how to behave, or that it would make very little difference if they heard because they would not understand, but his friends should try to see that Press has embarked on an enterprise of outstanding originality—that of attempting to influence how the US government spends dwindling resources in a good and durable cause. Plainly it would be tactless for one in Press's job to go about the government town in which he works complaining that the space station is a waste of money, and probably a shot in the foot of the US economy as well. But his friends are more free to say those shocking things.

The issue travels further than the United States. The US research enterprise has many faults. Even in basic science, competition is unnecessarily cut-throat. Too many people work too hard and think too little. Useful applications repeatedly fail to materialize because the trade-off between the vaulting ambition of researchers and the empiricism of those who would exploit them is rarely made openly. Yet the US research enterprise remains the best model for understanding the natural world. Would it not be folly to let it turn to sand because there is a financial crisis in an election year? □



Changes in environmental release laws possible

- OTA submits report to Congress
- Tiered risk assessment proposed

Washington

OPTIONS for the regulation of the deliberate release of genetically-engineered organisms were presented yesterday to the US Congress in a report* from the Office of Technology Assessment (OTA). Although the report avoids making specific recommendations, its conclusions suggest that a shake-up in the regulation of biotechnology is necessary.

Congressional interest in deliberate environmental release has been spurred by gaps and overlaps in the regulations of the various agencies given responsibility for biotechnology under the Coordinated Framework for the Regulation of Biotechnology. The Coordinated Framework, set up in 1986, is a patchwork of earlier regulations drawn from the Environmental Protection Agency (EPA), the Department of Agriculture, the National Institutes of Health and the Food and Drug Administration. Overall coordination is provided by the interagency Biotechnology Science Coordinating Committee (BSCC). But the agencies have yet to agree upon the definition of "genetically-engineered organism", and what is an environmental release.

Congressional efforts to force the agencies to tighten their regulations, and to confer statutory authority to the BSCC have failed so far. The current OTA report recommends the creation of another coordinating body with sufficient muscle to direct the preparation of review standards for field-tests among the various agencies.

Currently, proposals to field-test genetically-engineered organisms undergo case-by-case review by the appropriate agency — a process that is likely to be swamped once the number of biotechnology products increases. OTA suggests a stratified scheme whereby only organisms that contain genes from a pathogen or which pose a chance of disrupting the ecosystem into which they are introduced would be subject to rigorous review; other field-tests would be evaluated by comparing them to previous ones.

Adequate assessment of the risks posed by field-tests will depend upon better data on the movement of genetically-engineered organisms in the environment, according to the OTA study. Such research overlaps disciplinary boundaries, and would be encouraged by the creation of an interagency task force

or an increase in the level of funds for the study of natural selection, says OTA.

Despite the suggested changes, the OTA study concludes that "the adequate review of planned introductions is now possible". Even so, the report is more cautious than a similar assessment prepared by the National Academy of Sciences (see *Nature* 328, 653; 1987). Congress's reaction to the report will be gauged by whether the committees that requested it go on to submit legislation — a move that will probably be opposed by the biotechnology industry and the scientific community.

Carol Ezzell

*New Developments in Biotechnology—Field-Testing Engineered Organisms: Genetic and Ecological Issues. Office of Technology Assessment, Washington DC, 1988.

Acrimonious debate over AIDS

Washington

AIDS continues to arouse strong emotions in the US Senate, reflecting deep divisions within US society. Last week the Senate passed the "Federal AIDS Research Information and Care Act", co-sponsored by Senators Edward Kennedy (Democrat, Massachusetts) and Orrin Hatch (Republican, Utah) but not before the conservative Senator Jesse Helms (Republican, North Carolina) succeeded in adding a series of restrictive amendments.

The bill is intended to provide \$700 million for a comprehensive national plan to stem the spread of AIDS, for home and community care for AIDS patients, and for educational campaigns by both national and community organizations. Helms succeeded in adding amendments that barred any funds being spent on educational programmes that might condone homosexual behaviour. In a debate that was often angry, Helms said that he would "not allow one dollar of taxpayers' money to promote sodomy". He also succeeded in adding amendments that would make AIDS testing of sex and drug offenders compulsory and would ban plans to give clean needles to drug addicts.

In an attempt to reduce the damage caused by Helms, Kennedy succeeded in adding a further amendment to prevent the administration restricting "accurate information" about AIDS. The next step will be when the bill emerges from a subcommittee of the House of Representatives.

Alun Anderson

"Non" to man in space

Paris

THE European Space Agency's long-term plan for manned space expeditions has little scientific justification, according to the Space Research Committee of the French Academy of Sciences. In a report just published, the committee concludes that "the principal justification for manned flights envisaged by European nations is to affirm their ability to put men into space with purely European means".

Since the principal components of ESA's plan were approved by member states, apart from Britain, at the inter-ministerial meeting in The Hague last November, the Academy's report has come late in the day. But the findings should provide food for thought for ESA experts meeting in Strasbourg this week to discuss the manned missions. Ironically the two major elements in the ESA plan—the Hermes manned shuttle and the Ariane V launcher—are heavily backed by the French government.

ESA's science programme, which includes astronomy, Earth observation and microgravity experiments, will soon be capable of operating without the need for the presence of man. Indeed, such a presence "could be deleterious in the majority of space research operations". While some observations depend upon sensors in space, the report argues that remote operation from a ground station, or automatic data relaying are quite adequate. Furthermore, as satellites are often in orbits that are extremely hostile even for the well-protected astronaut, the role for man in the maintenance of satellites is also questionable.

Human intervention in some life science experiments aboard a space station will be necessary in the short term. For experiments on human physiology itself there is no alternative. But, says the report, such experiments do not, alone, justify European investment in manned space missions.

The Space Research Committee, whose report follows two years of study and debate in close liaison with the CNRS, the French national space research centre, feels that the desire to participate in the "human adventure" of manned space is "perfectly legitimate". What the committee fears—as the defender of purely scientific research—is that the high cost of putting man in space will take money away from basic science research "on the ground" in France. Such research, the committee argues, will be essential for the future success of Europe's space programme.

Peter Coles

Minister proposes 'Richter' scale for nuclear accidents

Paris

Two years after the explosion of the nuclear reactor at Chernobyl, the French Industry Minister, Alain Madelin, has drawn up a 'scale of severity' to evaluate nuclear accidents. In common with other European governments, the French have now learned that much of the reassuring information put out soon after the Chernobyl incident was at best misleading and at times false. This has undermined public confidence in the large French nuclear power programme, and Madelin's new approach is intended to help allay public fears by making it possible for the layman to appreciate the degrees of risk involved.

Madelin's scale has six levels, defined essentially in terms of the extent of pollution risk. Operational anomalies without immediate hazard are covered by the two lowest. Level 1 means that operating norms have been exceeded, while Level 2 incidents could have serious repercussions if no action is taken. The recent sodium leak in the Superphénix reactor

would have been Level 2.

Level 3 covers incidents where radioactive substances are leaked at concentrations over 10 per cent of the permitted annual threshold, or where containment structures are damaged. At Level 4, leaked radiation would reach authorized levels, or would imply medical attention for those affected. Also at this level are accidents to the reactor core itself.

At Level 5 and Level 6 the risk is no longer confined to the reactor site. The severity of the risk is defined by the level of leaked radiation—either tens of thousands of curies (Level 5) or hundreds of thousands of curies (Level 6).

Given the difficulties encountered, even within the European Community, in establishing what levels of radiation are 'safe', Madelin's scale has been criticized as giving a false image of objectivity. And the public is still at the mercy of a government's readiness to supply accurate and timely information in the event of a nuclear accident, however the information is expressed. **Peter Coles**

Atomic energy research 'under threat' from privatization

London

FURTHER warnings were sounded last week about the threat to research posed by the British government's plans to privatize the electricity supply industry. The United Kingdom Atomic Energy Authority, whose chief function is to carry out research and development in support of the nation's nuclear power programme, is particularly concerned that during the period leading up to the sell-off and when the industry is in private hands, cash for the fast reactor programme will dry up.

Giving evidence to a House of Lords select committee last week, the authority's chairman, John Collyear, said that uncertainties about the future responsibilities of a privately-owned electricity supply industry made planning difficult. Furthermore, he envisaged problems arising over the money that will be made available for underlying basic research. At present the authority's principal customers, the Department of Energy and the Central Electricity Generating Board (CEGB), allow 6 per cent of their spending to be used for basic research. The government proposes to sell off the CEGB in two lots—one retaining 70 per cent of generating capacity, including the nuclear power capability, and the other with 30 per cent. Collyear is concerned that both companies, especially the non-

nuclear company, would be reluctant to support the authority's work voluntarily.

Already the CEGB has pulled out of two authority-led research programmes, possibly because of impending privatization (see *Nature* 332, 382; 1988). The government's stated plans make no specific mention of research. The authority believes that the new structure of the industry should provide "fresh opportunities" for the authority to provide services. "But it is not likely that new entrants to the industry will be strong enough to sponsor research and development into longer-term advanced nuclear (or non-nuclear) technologies". The authority proposes funding through a levy on electricity sales or by subscriptions from companies.

Britain's principal centre for fast reactor research is at Dounreay in Scotland. The site operates a 250 MW prototype fast reactor, and an associated reprocessing plant. The reactor is unique among prototype fast reactors in having a facility for testing full-scale commercial fast reactor fuel, which can then be reprocessed at the nearby plant. The authority is waiting for permission from the government to sign a collaborative agreement with four European partners to build a full-scale demonstration reactor. **Simon Hadlington**

Indian ambitions

The Indian government is hoping to be invited to collaborate with the United States on the Superconducting Super Collider (SSC) project because, says Science Minister K. R. Narayanan, "participation in this experiment will benefit our country especially in the fields of high-energy particle physics and open up new areas in the frontiers of science".

Talks on Indian participation were held recently between Leon Lederman, director of Fermilab, and high-energy physicists from the Bhabha Atomic Research Centre and the Tata Institute of Fundamental Research in Bombay. Lederman said that a number of suggestions have been made but details have yet to be worked out. According to Lederman, one possible area of collaboration is in devices and equipment such as high-vacuum systems and superconducting correction coils.

Because of cheap labour, auxiliary equipment for the SSC could be made in India at low price. Indian physicists feel that the technology and experience they would gain in fabricating SSC components would be helpful in their own accelerator programme. **K. S. J.**

■ Although the US Department of Energy has made a general appeal to other countries to contribute to the SSC, both financially and in kind, collaboration on magnet building would require cooperation between industrial manufacturers as well as governments. Because one of the postulated benefits of SSC is to help American industry, the possibility of joint ventures in building SSC has been regarded with some ambivalence. **D.L.**

Canadian submarines

US President Ronald Reagan last week told visiting Canadian Prime Minister Brian Mulroney that he will allow Canada to use secret US nuclear submarine propulsion technology. Canada is considering buying a fleet of nuclear submarines from either Britain or France. Purchase of British "Trafalgar" nuclear attack submarines requires permission from the United States as they use US-designed power plants. President Reagan's decision is ironic in that Canada was prompted to buy submarines by a dispute with the United States over the sovereignty of the Arctic North-West passage. **A.A.**

Boyer/Cohen award

A NEW prize for discovery in 'pure or applied science in the field of biotechnology' has been endowed by two famous French names: Moët Hennessy and Louis Vuitton. The first prize goes to Herbert W. Boyer and Stanley N. Cohen for their pioneering work in genetic engineering. Valued at FF500,000 (£50,000), the award will be presented on 5 July at a conference at the Institut de la Vie, near Paris. **P.C.**

Prospects of marine wealth entice Japan's biotechnologists

Tokyo

JAPAN'S Ministry of International Trade and Industry (MITI) is looking to the sea to solve two seemingly unrelated problems—the rundown of the nation's heavy industries and foreign complaints that Japan's research system is closed to outsiders.

From October, MITI, in collaboration with shipbuilding, steel, construction, petrochemical and biotechnology companies, will start construction of two new marine research institutes as part of a ten-year project aimed at creating new industries for the extraction of 'fine chemicals' and other 'high-value-added' products from marine organisms. The birth of the new project accompanies a major reorganization of MITI-sponsored research which is intended to divert funds into the establishment of basic research institutes in the private sector that will be open to foreign researchers.

Under recently introduced legislation, the New Energy Development Organization (NEDO), set up jointly by MITI and private industry in 1980 to build pilot plants for tapping alternative energy resources, will be turned into the New Energy and Industrial Technology Development Organization. The new organization will have a much wider research and development role and will absorb several of MITI's big projects, such as the 'large-scale' and 'next-generation' industrial projects, as well as the new 'fine chemicals from marine organisms' project. It will also initiate two other new projects: a high-speed 'elevator' in a disused coalmine shaft for manufacturing new materials in a weightless environment, and an ion-beam engineering research institute.

The new NEDO will use funds provided by industry and by national and local government. Additional interest-free loans will be derived from the sale of government shares in the now partly privatized telecommunications giant, Nippon Telegraph and Telephone (NTT). Dividends from government-owned NTT shares are being used to support research centres such as MITI's Protein Engineering Research Institute but this is the first time that earnings from the NTT sale have gone into basic research.

The research institutes will be 'rented out' to users—mainly researchers from the companies that help to establish the institutes. But under the new NEDO framework, special government subsidies can be given to encourage the institutes to include foreigners in the research teams. MITI hopes that this step

will help to counter Western complaints that Japan's research system is closed.

The two marine research institutes will be built in Shimizu, near Tokyo, and in Kamaishi on the north-east coast of the main island of Honshu. The ports of Kamaishi and Shimizu were selected partly because they allow access to two major ocean currents—the cold Oyashio current and the warm Kuroshio current—and partly because the closure of heavy industrial plants in the two cities has provided land and harbour facilities. Nippon Steel recently closed its last blast furnace in Kamaishi, and Toa Nenryo Kogyo, a large oil refiner, has land available from a former plant in Shimizu. Both companies are key backers of the new institutes.

Six thousand million yen (about \$50 million) has been allocated for the institutes' construction in the second half of fiscal 1988 (which begins in October). And MITI hopes to pour another ¥15,000 million (\$120 million) into the project over the next ten years with a similar amount from private industry.

One of the participating companies, Suntory Ltd, a major distiller that has diversified into biotechnology, has been conducting marine research since 1979. Suntory has its own research vessel and

has been searching for new drugs, particularly anticancer and antihypertension agents, from corals collected from reefs in Australia, Palau and Okinawa. But MITI officials say they will not include research on drugs in their proposals to the Ministry of Finance, as this would encroach upon the territory of the Ministry of Health and Welfare. Similarly, fish research will remain in the domain of the Ministry of Agriculture, Forestry and Fisheries, but the project aims to develop 'fine chemicals' such as barnacle-resistant paints.

David Swinbanks

■ MITI is not the only ministry to bring 'biotechnology' projects to areas suffering from industrial decay. The Ministry of Agriculture, Forestry and Fisheries, under the auspices of BRAIN (Bio-oriented Technology Research Advancement Institution), another organization financed through the privatization of NTT, is developing aquaculture using tanks in a recently closed coal mine in Nagasaki prefecture. And the same organization, with the support of Nippon Steel and Hitachi Zosen, one of Japan's largest shipbuilding concerns, has set up the Iwate Biomass Institute to develop techniques to convert softwood from Japan's flagging forestry industry into feed for cattle. The mere mention of the word biotechnology sends local governments clamouring for these government projects, despite the fact that it will probably be years before they pay off in terms of new jobs.

Rock holds key to mineral riches

A FLEET of ships has set sail from Yokohama to save a vital piece of Japanese territory—two lumps of rock protruding less than a metre out of the south-west Pacific ocean. Okinotorishima, lying 6,800 km south of Tokyo, is a respectable little island of about five by two kilometres at low tide. But at high tide only two rocks are left sticking about 70 centimetres above the sea. If these are washed away Japan will lose about 400,000 square kilometres of its 200-mile exclusive economic zone and with it all rights to fishing and minerals in the area; the surrounding sea abounds with tuna and bonito, and cobalt-rich crusts have been found on the adjacent deep-sea floor.

Five years ago there were four rocks above high tide level. But erosion has reduced these to two propped up on fragile columns of rock. And as the island lies in the typhoon belt they could be swept away at any moment. The 22,000-ton *Sun-*

imosu Ace, mother ship of a four-vessel fleet, left Yokohama on April 22 laden with iron tetrapods, concrete, several small boats and two helicopters. Tetrapods will be laid around the two remain-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

High-tide at Okinotorishima (Kyodo Photo).

ing rocks of Okinotorishima to form a circular wall 50 metres in diameter, which will be reinforced with concrete. Automated meteorological observation equipment is expected to be installed on the island. Salvage of the "island" will take three years and cost about 30,000 million yen (\$240 million).

David Swinbanks

Pro-Solidarity group objects to governmental appointment

London

POLAND'S "Social Committee for Learning" (the SKN), an unofficial group which campaigns for academic freedom and provides stipends for young scholars and researchers excluded from the academic establishment on political grounds, has issued a strong protest against the appointment of Professor Benon Miskiewicz as chairman of the Central Qualifications Commission, which has the task of conferring senior academic degrees and titles.

Professor Miskiewicz, a former rector of Poznan University, and a military historian by training, was Minister of Science, Higher Education and Technology in 1982, during the period of martial law. He presided over the "pacification" of the universities with what the SKN calls "a brutal sincerity and an unswerving awareness that he was an exponent of the only correct line". Miskiewicz was therefore responsible for the dismissal of a number of academics on political grounds, including a former colleague, Professor Leszek Nowak, who held a tenured lectureship in philosophy at Poznan University and who resigned

from the Polish United Workers Party when martial law was declared.

In 1985, Professor Miskiewicz's ministry enacted amendments to the Higher Education Act which annulled the academic autonomy won during the Solidarity period—although a nationwide referendum had revealed almost unanimous opposition of the academic community to these changes. Miskiewicz lost his ministerial post last October, when Poland's two education ministries were merged.

The appointment of someone with such a background is, says the SKN, "an ominous and provocative decision".

For obvious reasons, the SKN does not publish the names of its members, but it is known to include several members of the Polish Academy of Sciences and senior university personnel. This latest protest coincides with a new surge of political unrest in the universities, with the staff campaigning for the restoration of the banned Solidarity-style union groups, and the students demanding the legalization of the Independent Students' Association of the Solidarity period.

Vera Rich

Relief in sight for HIV-carrying Japanese haemophiliacs

Tokyo

SOME relief may be on the way for Japanese haemophiliacs infected by blood coagulants contaminated with human immunodeficiency virus (HIV), the virus that causes AIDS. About 40 per cent of Japan's haemophiliacs, or 2,000 people, are believed to have been infected with HIV through blood products imported from the United States.

Japan was slow to introduce heat treatment of coagulants (blood-clotting factors) to kill the virus, but the Health and Welfare Ministry denies direct responsibility and refuses to recognize HIV infection as the side-effect of a drug. As a result, haemophiliacs who are asymptomatic carriers of HIV cannot get full national health insurance cover.

Prime Minister Noboru Takeshita has promised government help. In response, the LDP subcommittee on AIDS, following a proposal from the Ministry of Health and Welfare, has recommended that Japan be divided into eight or nine regional blocks in each of which key university and national hospitals will be asked to act as bases for the treatment and counselling of haemophiliacs with AIDS problems.

Government money will be used to

cover the costs of treating asymptomatic patients and there will be measures to cover the extra bed costs of hospitalized haemophiliacs suffering from AIDS. But haemophiliacs are not satisfied. The National Association of Haemophiliacs has submitted a written demand to the Health and Welfare Ministry for "complete relief measures" from the government, including adequate compensation and the best possible treatment for those infected.

A Kyoto-based group of "victims of imported blood products" is demanding compensation for the families of victims who have already died, and "sympathy money" and annuities for living expenses for those with AIDS.

The ruling party is also hoping to reopen deliberations on the controversial new 'AIDS Prevention Law' which has been stalled in the Diet since last year by opposition from lawyers, physicians and haemophiliacs who say the law will violate patients' human rights. Yoshiaki Ishida, for the Kyoto group, says the subcommittee's recommendations are a step in the right direction, but if the proposals are tied to acceptance of the new AIDS law, they are "nothing to be happy about".

David Swinbanks

Australian fracas

MOUNT Etna, in Rockhampton in Queensland, has become the scene of a bizarre confrontation between environmentalists and the owners of the mountain, Central Queensland Cement. Mount Etna provides the company with limestone worth some A\$18 million annually. But now Central cannot carry out the blasting needed to extract more limestone. Conservationists, led by Mr Craig Hardy, have hidden themselves in the vast warren of caves within the mountain. They say they will stay there. They aim to save the 46 named caves and a rare colony of bats, and have food supplies to last months.

"Hundreds of people are telling us to blow them up. But we're a responsible company", explained Central's general manager, Mr Robin Town. Time may be on the company's side, though, because there are stockpiles of enough limestone to last 5 years. Untapped reserves are estimated at 30 years' supply.

T.E.

FRG/China deal

A NEW agreement will make it easier for researchers in West Germany and the People's Republic of China to collaborate. It was signed by representatives of the Deutsche Forschungsgemeinschaft (DFG) and the National Natural Science Foundation (NSFC) in Peking on 25 March. The agreement complements other pacts between the two countries. Over 500 researchers have gone to China from Germany since 1974 under the auspices of the Max Planck Society and the Chinese Academy of Sciences. The DFG made a similar agreement with the Chinese Academy of Social Sciences in 1981.

S.D.

Universities neglected

CHINA is devoting too few resources to too many universities, according to Professor Ding Shishun, president of Beijing University. "Some leading comrades" pay attention only to building and running factories, and consider education a "soft" task, he said. Instead, Professor Ding called for China to follow the example of postwar Japan and West Germany, which invested in education in spite of severe economic problems.

V.R.

AIDS vaccine first

ENCOURAGING results have come from the United States' first human trials of an AIDS vaccine. Joseph Kovacs of the National Institutes of Health and colleagues reported at the American Federation for Clinical Research's meeting that six of 15 volunteers who received a 40 microgram dose of recombinant gp160, the envelope protein of the human immunodeficiency virus, showed an antibody response to gp160 after eight weeks. Higher doses of gp160 and the nature of the antibody response will now be investigated.

A.A.

Arizona team spins its biggest telescope mirror so far

Tucson

THE giant mirror-making furnace at the University of Arizona took another spin on 25 April, and the first of two 3.5-metre telescope mirrors came out glowing. This was the second trial for the Steward Observatory Mirror Laboratory's rotating furnace, which its designers say should make mirror-making less expensive and less time-consuming.

Jan McCoy

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

More than two tons of glass is loaded into the mould.

The honeycomb sandwich design results in a stiff, light mirror which can more easily be kept close to air temperature. This helps keep images crisp, according to J. Roger Angel, the University of Arizona astronomer who came up with the mirror design.

For this casting, it took 1.5 days to load the 5,000 pounds of borosilicate glass into the 6-metre furnace. Two hun-

dred kilowatts were used to heat the glass to about 1,170°C, and another 130 kilowatts were needed to keep it at that temperature. After about two hours at the peak temperature, the furnace was allowed to cool. After 27 hours, with the temperature below 600°C, the spinning was stopped. Slow cooling will continue for six weeks to allow the glass to anneal. Next year the furnace will be expanded to 10 metres for making a 6.5-metre mirror to upgrade the Multiple Mirror Telescope on Mount Hopkins in Arizona.

This first 3.5-metre mirror will be named after Marc A. Aaronson, the University of Arizona astronomer killed in an accident on Kitt Peak on 30 April 1987. It will be used in the Astrophysical Research Consortium telescope on Apache Point in New Mexico. That telescope will be the world's 14th largest and will be operated remotely by the five universities in the consortium, Princeton, the University of Chicago, New Mexico State, the University of Washington and Washington State.

Elizabeth Pennisi

■ France has signed an agreement to cooperate with Spain in the field of astrophysics, thus joining Denmark, Great Britain, Sweden and West Germany. The Themis Solar Telescope, developed by the French National Institute for Sciences of the Universe is to begin construction on the island of Tenerife in the Canaries next spring.

P.C.

US astronomers have to choose which telescopes to keep

Washington

ASTRONOMERS in the United States, their funding eroded by inflation over the past decade, now find themselves forced to choose between the maintenance of existing telescope facilities and the planning and construction of new ones. Although no hard decisions have yet been made, it seems inevitable that some of the smaller facilities at notable observatories such as Kitt Peak in Arizona and Cerro Tololo in Chile will be curtailed.

Kitt Peak and Cerro Tololo are run by the National Optical Astronomy Observatories (NOAO), which is operated by the Association of Universities for Research in Astronomy (AURA) and supported financially by the National Science Foundation (NSF). NSF has given approval for several new projects, notably the Global Oscillations Network (GONG), a worldwide chain of solar

telescopes, but has given NOAO no extra money to support them. This means that modest budget items, such as paying journal page-charges for visiting astronomers, must now be sacrificed to keep the telescopes working.

In a letter to the US astronomical community earlier this year, Goetz Ortel, president of AURA, and Sidney Wolff, director of NOAO, described three strategies for survival: maintain existing operations at the expense of new projects; abandon all but the biggest present telescopes and concentrate on building for the future; or support only those facilities where NOAO can make a unique contribution to world astronomy. But even if the first plan were adopted, it is not clear that smaller telescopes such as the 1.5-metre at Kitt Peak would remain in operation.

Ortel says that the community has

All go for Canada and space station

Edmonton, Alberta

CANADA's space ambitions took one step forward and one step back last month. On the plus side, the government has resolved the financial question raised by the extra Cdn\$388 million needed to pay for Canada's role in the design, development operation and use of the international space station being organized by the US National Aeronautics and Space Administration (NASA). But at the same time, the British decision not to participate in Radarsat, a remote sensing satellite (*Nature* 332, 669; 1988), has sent that project into limbo.

Canadian participation in the space station became a hot political question earlier this year when the estimated cost of participating climbed from Cdn\$800 million to nearly Cdn\$1,200 million. But Canada sees space station participation as crucial for sustaining future economic growth, and it seemed inevitable that the money would be found. Robert de Cotret minister of state for science and technology hailed the funding decision, calling space station participation "a measure of our technology maturity as a nation".

Enthusiasm in the scientific community was tempered by the fact that part of the additional funds needed will come out of a Cdn\$1,300 million 5-year programme announced earlier this year to support science and technology. At a conference here entitled "University research and the future of Canada" there were rumblings of discontent as financially strapped research scientists wondered how much of the new money would find its way to the Canadian granting councils, the primary source of funds for basic research. Without British support, Radarsat faces an uncertain future. Although the project is "far from dead" according to a government spokeswoman, she acknowledged that the prospects for finding a new partner are not bright.

Joseph Palca

reacted constructively to the crisis, and he has received thoughtful and imaginative responses from several astronomers. One suggestion is that universities could lease or take over operations which NOAO can no longer afford.

An underlying concern to the US community is that European astronomers are regaining an ascendancy they lost about fifteen years ago. Jacques Beckers, who was in charge of NOAO's National New Technology Telescope programme, has left to direct the corresponding European venture, the Very Large Telescope, which has been given financial support. If US astronomers are to compete in the future, says Ortel, painful choices must be made now.

David Lindley

UK scientists must take lead in steering inevitable changes

London

TRADITIONAL methods of organizing and financing British academic science can no longer support a system that has ceased to expand, and the scientists themselves should play a significant role in arriving at alternative schemes. Such were the two chief conclusions from a conference last week organized by the British Association for the Advancement of Science and aimed at prodding Britain's research community into having some input in determining its own fate. The title of the conference, 'Managing Science in a Steady State', was taken from a report of the Science Policy Research Group, an independent unit set up and supported by the research councils.

The essential argument of the report's author, John Ziman, is that because further expansion of scientific activity will be limited by economic and political considerations, a structure and management system must be developed that will allow adjustment to change within a constant envelope of resources, while being expected to serve the nation more efficiently and account more directly for costs. "The traditional academic organization of basic science evolved under conditions of expansion", Ziman told last week's meeting. "There is no way back to traditional academic arrangements for science. To attempt it would be trying to kick history in the face and that would really be the end of British academic science."

While there was general consensus on Ziman's analysis, suggestions about how

to arrive at solutions were more contentious. John Irvine, of the Science Policy Research Unit at Sussex University, was politely if not enthusiastically received for his contribution on the growing validity and acceptance of research evaluation techniques, including bibliometric analysis.

Funding agencies required a change in philosophy, Irvine argued, from regarding assessment techniques as *post-hoc* auditing to accepting them as "an integral part of the research management system". Inevitably, the discussion turned to the sensitive topic of the subject review programme currently being carried out by the University Grants Committee. With the earth sciences review now complete, physics and chemistry are being examined. K. Edwards, of the University of Leicester, pointed out that the outcome of previous reviews is likely to have an increasingly marked influence on successive reviews. An institution whose chemistry and physics departments had been forced to close would be unlikely to be able to support a biology programme.

Given that the government is believed to be formulating proposals for the future shape of academic science and is expected to make an announcement later in the year, last week's meeting (with some 150 delegates, including several from the Cabinet Office) could turn out to be the last major forum for debate before major decisions are taken. If the scientists want to have a say, they should act quickly.

Simon Hadlington

Sweet words on acid rain

Washington

CANADIAN Prime Minister Brian Mulroney ended his official visit to the United States with a promise from US President Ronald Reagan that an agreement to reduce acid rain would be "taken up as a matter of priority". Although Mulroney described Reagan's words as "encouraging", the prospects for a rapid change in US policy still seem slim.

Acid rain is the biggest cause of friction between the two nations. The Canadian side argues that more than half and, in places, as much as seventy per cent of the acid rain falling on eastern Canada comes from the burning of fossil fuels in the United States.

Mulroney presented Reagan with an eight-point proposal for an acid rain treaty. But the reaction from Reagan appeared similar to that given in April 1987 when he addressed the Canadian parliament and promised to "consider the Prime Minister's proposal for a bilateral accord on acid rain". No progress has been reported since then. At a press conference after his final meeting with Reagan, Mulroney answered complaints that he had obtained no clear commitment to reduce acid rain by saying "What do you do? Do you declare war? Do you interrupt relations? Or do you actively work to persuade Americans?". Mulroney believes that although "Canada has been disappointed before", an acid rain treaty between the United States and Canada will come "as surely as summer follows spring".

Canada has acted to remove its own sources of acid rain with a control programme that will reduce sulphur dioxide emissions in eastern Canada to half of 1980 levels by 1994.

Mulroney's visit coincided with the release of a new report from the Washington-based Environmental Defense Fund showing that nitrates deposited in acid rain are a major source of pollution to eastern US coastal waters. Nitrates cause algal blooms which reduce light and oxygen levels.

US power companies—the major consumers of fossil fuels—continue to argue through their association, the Edison Electric Institute, that "rain acidity at current levels does not threaten the environment". They estimate that the cost of a major reduction in sulphur and nitrate emission would be a prohibitive \$6,200 million annually. Instead, they argue that they should be left to invest in new clean coal technologies which could be ready for the late 1990s when many of the current plants will be due for replacement.

Alun Anderson

Who will lead human genome project?

Washington

CONTROVERSY over who will be the lead agency in the human genome project erupted again last week as James Wynn-gaarden from the National Institutes of Health (NIH) and David Nelson of the Department of Energy (DOE) testified before a subcommittee of the US House of Representatives Committee on Energy and Commerce.

Questions posed by the subcommittee followed the line taken by the Office of Technology Assessment's report on the project (see *Nature* 332, 769; 1988) and indicated that NIH and DOE will share in leadership. That flies in the face of advice given by the National Research Council, which specified the need for a lead agency in order to reduce the time and money spent on coordination between organizations.

During the hearing, DOE indicated

their desire to share in the project rather than take control while NIH leant more towards establishing a single lead agency. Both organizations are very committed to the project but others will be involved. The National Science Foundation could be a source of revenue for instrumentation development.

Legislation to decide leadership could materialize from the Committee on Energy and Commerce after requested figures on budget breakdowns in the area of technology development arrive from the DOE. In the Senate there is related legislation in the form of a bill which carries a provision for the development of a National Advisory Panel on the Human Genome, and it is hoped that the bill will be passed through the Senate by the end of May. With the multitude of issues facing the 100th Congress, though, this is unlikely.

Elizabeth Ebbert

Electrons collide at Stanford

Berkeley

AFTER a year of delays due to budget cuts and technical difficulties, Stanford's Linear Collider (SLC) should soon be producing Z^0 particles, carriers of the nuclear weak force. On 18 April, beams of electrons and positrons 16 micrometres in diameter were for the first time made to pass through each other in the collider's detector hall. Combined beam energies of 93 GeV, enough for Z^0 production, were achieved, and within 2 months scientists at the Stanford Linear Accelerator Center (SLAC) expect to begin experiments.

The SLC, which was built for \$115 million, is the ultimate refinement of high-energy linear electron-positron colliders, and is a rival to conventional storage rings, which become prohibitively expensive at energies of greater than 100 GeV. It uses Stanford's linear accelerator to accelerate bunches of positrons and electrons, which are then guided around opposing arcs that focus them on a collision course with each other.

Gramm-Rudman budget cuts last year slowed the commissioning process, as did unexpected difficulties due to broadening of the particle beams as they travel through the collider arcs. The problem was hard to correct because of the complex geometry of the arcs, which for reasons of economy do not lie in a plane but were built to follow the contours of the hillside. The beams were finally condensed and focused by adjustment of about half of the 500 guiding magnets.

Before experiments can begin, the beam intensity must be doubled, and adjustments may also be necessary to minimize background radiation, but neither task represents a major hurdle, said SLAC spokesman Michael Riordan. The collider is expected to yield only a few Z^0 particles a day initially, and an optimal beam diameter of 3 micrometres, yielding 3,000 Z^0 s per day, may not be reached for several years.

If it meets no further snags, SLC should have the first chance to determine the mass, lifetime and decay modes of the Z^0 . More detailed study of it is likely to be done at CERN's large electron-positron collider (LEP), expected to produce Z^0 particles at five times SLC's optimum rate.

Marcia Barinaga

Glaxo award

Nature's editor John Maddox has been awarded the Glaxo Science Writers' Award for "Best Article or Series of Articles in a Specialist Journal on a Science Subject". The award was given for the "Science in the Soviet Union" supplement in *Nature* 29 October 1987.

Anti-nuclear emotions raised by Bavarian reprocessing plant

Munich

MARATHONS, balloons, films, and violent and non-violent protests are only some of the means that West German anti-nuclear activists have used to demonstrate their opposition to the Wackersdorf nuclear reprocessing plant. And last week, opponents presented the Bavarian Environment Ministry with petitions containing over 700,000 signatures gathered in West Germany, Austria and even the United States, calling for a halt in construction of the plant.

Two legal challenges being considered by the Federal Constitutional Court may prove more effective than all the other protests combined in shutting down Wackersdorf. The plant is due to begin operation in 1996 at an estimated final cost of DM 7,400 million.

The petitions are likely to present logistical problems for the Bavarian Environment Ministry. The ministry is required to hold a public meeting for opponents, licensing authorities and DWK, the firm constructing Wackersdorf. If even 10 per cent of the opponents were to come, the meeting might have to be held in Munich's Olympic Stadium.

One legal challenge might give the opponents a chance to stop the construction of Wackersdorf by referendum. Bavaria's highest court had rejected a referendum on the location of the plant, declaring that the decision rests with the federal government. The opponents claim the Bavarian court contained too many members of a single political party, the Christian Social Union, which generally supports the nuclear programme.

The second case before the federal court concerns a January decision by a Bavarian court that declared the development scheme for Wackersdorf null and void because of insufficient protection of the population against radiation in the event of an accident. Construction has continued, however, because the Bavarian Interior Ministry ruled that the individual construction licences for sections of the plant remain valid even if the overall scheme is invalidated.

Wackersdorf opponents also challenge the decision by another federal court to allow construction on the basis of the construction licences, despite the lack of the necessary "nuclear licence" for the plant. Opposition leader Armin Weiss (Green Party), a chemistry professor at the University of Munich and a member of the Bavarian parliament, denounced the inconsistencies in the rulings as "Orwellian doublethink."

Weiss questions the logic behind building a reprocessing plant when it

appears West Germany may never complete its nuclear fuel cycle. Reprocessing nuclear fuel elements is more expensive and less safe than putting spent fuel rods directly into "permanent storage sites", he said. The only reason to put Wackersdorf into operation, Weiss said, would be as a source of plutonium for nuclear weapons. (West Germany is prevented from possessing such weapons by the international nuclear Non-Proliferation Treaty, which lapses in 1995.)

The West German government stands by its nuclear programme. Marlene Mühe at the Environment Ministry re-

Encouragement needed

A LACK of motivation is prevalent at East German universities, especially among women in technical fields. This is the conclusion of a survey conducted at the Technical University of Dresden, reported in the magazine *Für Dich*.

Of university students in East Germany, 54 per cent are women. Even at the Technical University, 40 per cent of students are women, compared with close to nil in the 1950s. They achieve their hard-won places through good grades, but the survey reports that women entering food chemistry or textile technology have only a vague grasp of these fields.

Education researchers find the root of the problem lies in girls being actively discouraged in their interest in technical areas. Schools have succeeded in demystifying the study of mathematics (*Für Dich* reports that mathematics is the favourite subject of most girls as well as boys), but have done little to encourage women in technical fields.

Aware of the technological challenge presented by the gradual opening of East Germany to the West (see *Nature* 329, 192; and 330, 510; 1987), the East German Communist Party addressed the issue at its eleventh party congress, reminding schools that they are polytechnic in nature and should behave accordingly.

Steven Dickman

jected Weiss' weapons charge as "absolute nonsense." Environment Ministry official Armin Hagen listed several justifications for Wackersdorf. He said that even without a fast breeder, it makes sense to reprocess spent fuel elements because it saves energy. Furthermore, the reprocessing of fuel elements would save space in permanent storage sites.

But the last reason on Hagen's list for construction to continue is most compelling politically: stopping in mid-1988 would be like "throwing DM 2,000 million out the window".

Steven Dickman

Newton's "claw", not paw!

SIR—Did Newton discover Darwin (*Nature* 332, 200; 1988)? Was the *Origin* anticipated by the *Principia*, the *Descent of Man* derived from the dispute with Leibniz, and evolution just a sort of flawed recoinage? Were Darwin's paw and Newton's claw equally the marks of genius? I am puzzled to find Darwin everywhere in Nature.

DICK LINDENBAUM

Department of Medical Genetics,
Churchill Hospital,
Oxford, OX3 7LJ, UK

SIR—Physical scientists have long known that your editorial bias is to biology, and will so account for your editorial headline "Darwin's 'claw', not 'paw'"! To John Faulkner's letter, which has nothing to do with Darwin but is about Newton's solution of the brachistochrone problem. Brachisto-, (*Greek*, shortest) please, not brachystro-.

These confusions removed, the puzzle now is to understand the source of Faulkner's version of the remark attributed to Bernoulli, which he supposes to have been "passed down to generations of British schoolboys" as *I recognize the lion by his paw*. As a schoolboy, or perhaps undergraduate, I remember it being taught to me as *by his claw* and the vivid phrase has remained scratched into my mind for 40 years. This was the accepted translation in most of the earlier English language sources, for example in Sir David Brewster's standard two-volume *Memoirs of the Life . . . of Sir Isaac Newton* (Edinburgh, 1855), which was widely consulted by later writers.

The earliest use of "paw" I have discovered is in E.T. Bell's lively but not always accurate *Men of Mathematics* (London, 1937): "On seeing the solution Bernoulli at once exclaimed, 'Ah, I recognize the lion by his paw.' (This is not an exact translation of Bernoulli's Latin)." As Faulkner says, "claw" is the most straightforward translation: Bell's parenthetical observation suggests to me that he was relying on his memory when writing, rather than censoring the more convoluted alternative that Faulkner proposes.

DAVID W. DEWHIRST

Institute of Astronomy,
Madingley Road,
Cambridge CB30HA, UK

SIR—Bernoulli's famous lion-claw remark has its origin in a much older saying, regarding the immortal Phidias, whose acute sense of proportion is reputed to have enabled him to tell the size of a lion merely from its paw! Paw or claw, it matters not which, Newton's *griffe* revealed not the identity, but the magni-

tude, of its bearer — and so gave the game away.

DAVID ABOAV

29 Clements Road,
Chorleywood,
Hertfordshire WD3 5JS

1. D.T. Whiteside, *The mathematical papers of Isaac Newton*. VIII, 9N. (Cambridge University Press, 1981).

Risk research

SIR—Having been involved in research on the effects of 50 Hz electric fields¹, I fully agree with Foster and Pickard² about the problems in research on biological effects of electromagnetic fields, and, I would add more generally, of industrial agents. I wonder, nevertheless, whether Foster and Pickard do not miss one of the causes of the ambiguity they denounce, namely the ambivalence of reviewers and editors of major scientific journals when confronted with a manuscript dealing with possible dangers of industrial agents, an ambivalence arising perhaps from a more or less conscious fear of being accused of selling out to international capitalism. Libel laws being what they are, it might be awkward to cite chapter and verse, but it seems to me evident that some of the "positive" papers cited in the review by Carstensen³ would not normally have been accepted in the respected journals in which they appeared. Particularly glaring is often the inadequate description of methods, as noted by Foster and Pickard concerning their reference 11, which prevents close replication of experiments and which should have sown a healthy scepticism in the minds of reviewers and editors alike.

The guidelines needed for risk research are simple — scientific journals should enforce identical standards for "risk research" papers and ordinary research papers.

JANINE PERRE

Laboratoire Montyon-La Salpêtrière,
47 Boulevard de L'Hopital,
75651 Paris Cedex 13,
France

1. Portet, R., Cahannes, J., Perre, J. & Delost, H. *C. r. Soc. Biol. (Paris)* 178, 142–152 (1984).
2. Foster, K.R. & Pickard, W.F. *Nature* 33, 531–532 (1987).
3. Carstensen, E.L. *Biological Effects of Transmission Line Fields* (Elsevier, New York, 1987).

UK science teaching

SIR—Science teachers in Britain are being pressed to introduce changes that pose a serious threat to standards, especially for pupils of high academic ability. I refer to what is variously called "integrated science", "balanced science", or simply "science". The changes have apparently been approved by scientific and profes-

sional bodies in the hope that they will increase the proportion of able pupils who choose a scientific career, but it is hard to believe that their implications have been fully understood.

Two features of the proposals give reason for disquiet. The first is that integrated science will take up at most two subjects in the GCSE curriculum. Pupils who at present take physics and chemistry separately will therefore be presented with a diluted version of those subjects because one-third less time will be allotted; further dilution will occur because the syllabus is necessarily aimed at pupils who have considerably less ability to handle mathematical topics.

The second source of anxiety is the pressure upon specialist teachers of physics, chemistry and biology to teach outside their own specialities. Is it an appropriate use of scarce manpower for physicists to labour over the microscopy of living cells, or for biologists to struggle with the concept of valency? What will the consequences be for recruitment if science teachers are to be prevented from teaching what they know and enjoy and understand best? And have the implications for safety in school science laboratories received any attention from the reformers?

A knock-on effect upon university standards appears to be inevitable. Advanced-level physical science syllabuses are already being down-graded to take account of the change from O-level to GCSE; if a majority of pupils take two-subject "science", further downward revision can be confidently predicted. How much longer will British universities be able to teach to the standard of a professional scientific qualification in only three years? And what chance will there be of broadening our notorious sixth-form curriculum?

M.H. DODSON

High Lawns,
West Chevin Road,
Otley,
West Yorkshire LS21 3DL, UK

Lost axion

SIR—Surely your question of what has happened to the axion (*Nature* 331, 385; 1988) is easily answered?

I am assured by a friendly professor of nuclear physics that when *le bon Dieu* observes that he and his colleagues are stuck, he creates a new particle and sets it down just in front of where they are working, so that they may joyfully discover the New Creation and be spurred on. All that has happened with the axion is that God has had second thoughts and cashiered the thing, though perhaps only temporarily.

P.C. SMETHURST

89 Apton Road,
Bishop's Stortford,
Hertfordshire CM23 3ST, UK

Eye of the beholder

SIR—I applaud your decision to publish the letter from R.T. Chelvam (*Nature* 331, 10; 1988), as it both gives “scientific creationists” a forum so that they might not claim their ideas are suppressed, and exposes their outrageous scientific prevocations and dubious theology to scrutiny.

In support of his contention that “. . . it takes more gullibility to believe in darwinism than in Genesis”, Chelvam quotes Darwin¹ out of context thus: “to suppose that the eye . . . could have been formed by natural selection, seems, I freely confess, absurd in the highest degree”. Darwin’s unquoted, following sentence is, “Yet, . . . if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist . . . and if any variation . . . in the organ be ever useful to an animal order changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection . . . can hardly be considered real.”

L. v. Salvini-Plawen and Ernst Mayr² have displayed two such graded series of eyes in living prosobranch gastropods and polychaete worms. The body of evidence supporting “creation science” is thus smaller than Chelvam would have us believe.

Chelvam clearly proposes to hold the Creator personally responsible for all the details of biological design, without appreciating the theological problems that result. A notable example is again the human (or any vertebrate) eye, which is so badly bungled that a mortal engineer would be defenceless before a product liability suit. The retina is installed backwards, with the neural connection to the brain interposed between the photoreceptor and the light source. As a result, the neural connections must eventually be gathered together and brought through the retina, resulting in the well-known blind spot that the reader will perceive lateral to the fovea.

This is no trivial defect, but the inevitable product of an evolutionary process constrained at an early stage, when the orientation of the photoreceptors in a simple eyespot was of no significance. Rather than support the concept of an incompetent Creator that follows from biblical literalism, thoughtful theologians see Scripture and the Creator more as did Isaac Newton in his 1681 letter to Thomas Burnett: “As to Moses . . . he described realities in a language artificially adapted to ye sense of ye vulgar. . . . Where natural causes are at hand, God uses them as instruments in his works, but I do not think them sufficient for ye creation. . . .”

Evolution can be seen as one of the

instruments used in the creation. This doctrine raises fewer theological difficulties than does Chelvam’s biblical literalism. The initial legal opposition to “scientific creationism” came more from the US Christian community than from the scientific community³. I refer the reader to Conrad Hyers’ readable book⁴ for further exposition of the theological issues.

GEORGE S. BAKKEN

Department of Life Sciences,
Indiana State University,
Terre Haute, Indiana 47809, USA

1. Darwin, C. *On the Origin of Species*. (A facsimile of the First Edition with an Introduction by Ernst Mayr) 186–187 (Harvard, Cambridge, Massachusetts, 1964).
2. Salvini-Plawen, L. v. & Mayr, E. in *Evolutionary Biology* (eds Hecht, M. K. et al.) 10, 207–263 (Plenum, New York, 1977).
3. Overton, W. R. *Science* 215, 934–943 (1982).
4. Hyers, C. *The Meaning of Creation: Genesis and Modern Science* (Knox, Atlanta, Georgia, 1984).

First sentences

SIR—John Maddox¹ is no doubt right about the importance of first sentences. I had what I thought was a good one for a Letter to *Nature*²: “Consider an early Precambrian sea on a summer’s day”. But it was relegated to second sentence by the editors.

A. G. CAIRNS-SMITH

Chemistry Department,
The University,
Glasgow G12 8QQ, UK

1. *Nature* 332, 391 (1988).
2. Cairns-Smith, A. G. *Nature* 276, 807–808 (1978).

Own goal?

SIR—As the spread of the AIDS virus (human immune deficiency virus, HIV) has reached epidemic proportions in some communities, there has been much speculation on how this has happened. A number of sources, including your leading article (*Nature* 331, 376; 1988), have been trying to tell us that sexual promiscuity has had nothing to do with it. I refer to your criticism of Princess Anne’s statement that AIDS is an “own goal for mankind”. The spread of this virus has been inextricably linked to sexual promiscuity and together with intravenous drug-taking is recognized as one of the two major “amplification” factors clearly responsible for the spread of HIV. The extreme promiscuity of certain groups will lead to an exponential increase in the number of contacts, for example prostitutes in some Central African states and American west coast homosexuals, many of whom claimed that 1,000–2,000 partners a year was not uncommon. The more mobile and expanding global population of recent decades has also enabled a moderately infectious virus present originally at a very low level

in the human population to spread rapidly between separated populations of high risk individuals.

The morality and psychology of promiscuous behaviour is a matter for another discussion. However, it is most likely that if mankind had been monogamous the AIDS epidemic would never have occurred. The question Dr Gallo could have been answering was when?, rather than why now? Certainly the way forward is not to apportion blame but to understand and eradicate the disease. But to pretend that promiscuity is irrelevant to this epidemic is dangerously misleading.

FERGUS DAVISON

Liver Unit,
King’s College Hospital,
London SE5 8RX, UK

Flight recorders

SIR—In the News article “New information throws light on how Yuri Gagarin died” (*Nature* 331, 292; 1988), the admission by Professor Belotsekovskii that MiG15s did not carry accident data recorders in 1968 was described as “surprising”. In fact it would have been very surprising indeed if the Soviet Union had fitted accident data recorders to their fighter trainer aircraft in 1968, especially as a retrofit. The MiG15 design dates back to the late 1940s.

In the United Kingdom, the accident data recorder ‘came of age’ in 1965 when a British European Airways Vanguard aircraft crashed at London Heathrow Airport. This was one of only two out of the fleet of twenty aircraft to carry a recorder. The recording was found, by the public inquiry, to provide valuable evidence as to the cause of the crash. As from mid-1966, all new civil British-operated aircraft were required by law to provide ‘non-destructible’ records of flight data for accident investigation purposes.

Accident data recorders were thus originally introduced so that lessons learnt from accident data could help prevent other tragedies. As the systems have developed, it has become possible quickly to access the data for aircraft maintenance scheduling, providing considerable cost savings over regular maintenance based on flight hours alone. It has thus become cost-effective for military aircraft, but it is only in the 1980s that an accident data recorder has become a standard requirement for new British fighters and trainers.

Assuming that similar economic imperatives pertain in other countries operating two-seat fighters or trainers, it would be surprising if any such aircraft has had an accident data recorder as standard fit before 1980.

T.J. SPRIGGS

Island Close,
Hayling Island,
Hampshire, UK

Finding wood among the trees

Is the reductionist ambition for molecular biology in danger of being thwarted by the volume of the data it produces, or even by the absorbing interest of its collection?

Is molecular biology running into a dead end? Is it possible that the exciting programme to make biology a branch of physical chemistry is at risk of frustration by the sheer volume of the data now being acquired, by the complexity of some of the problems whose solution seems essential for the success of the reductionist programme — and by the honest fun of manipulating the techniques now available?

It is a curious situation, perhaps engendered by the plethora of discovery in the past thirty years and more. The genetic code, messenger RNA and its function, reverse transcriptase and (quickly) the techniques of genetic manipulation came treading on each other's heels. Nobody can cavil at what has been accomplished.

Most vividly, it would not have been possible to form a view of the cause of AIDS had it not been for the previous quarter of a century's molecular biology. More important, people have a sense that things will eventually be understood. Hardly a week passes without the publication of the amino-acid sequence of yet another cell-membrane protein, usually a receptor of some kind. Protein-DNA interactions, generally believed to determine the regulation of genes, are being dissected. There are now seven interleukins. People are also describing in previously unimagined detail the means by which information is relayed within cells, from membrane to nucleus and in the other direction, as in the processing of proteins *en route* from ribosomes to effector sites.

But now, while nobody can fail to marvel at the variety of the data accumulated, the volume of it is its most impressive attribute. The databanks storing nucleotide and protein sequences would be even further behind if laboratory people were as diligent as they should be in communicating their sequences quickly. Soon there will be comparable problems in storing authentic data from structural studies of biomolecules. If there is a cause for disappointment, it is that the reductionist prospectus should so slowly deliver its promises.

Naturally, it would be folly to ask that biology should become a branch of mathematics; physical chemistry is a sufficiently tall order. Even in the physical sciences, the explanations derived from first principles have not driven out experiment; the whole of hydrodynamics has not been reduced to the solution

of the Navier-Stokes equations. But, outside particle physics and cosmology, there is a sense in which the reductionist recipe works. Students, for example, can be taught the principles and helped to understand when and why they will not suffice.

Sadly, it remains a proper question of molecular biology whether the recent torrent of description amounts to understanding. Future historians may think it odd that so much should have been learned about the molecules on which life depends while so little has been understood about their function, or about life itself. The obvious parallel is with spectroscopy in the 1920s: technique had made it possible to fill catalogues with the frequencies of spectral lines, the search for empirical explanations was vigorously conducted, but only the coming of quantum mechanics made sense of all the data. The obvious answer is that the parallel may be false, and that biology has necessarily complicated explanations.

An example is the problem of protein folding. Should it not by now be almost a problem for undergraduates to solve, to ask what will be the conformation (tertiary structure) of a protein whose amino-acid sequence is known, or has been specified? That, alas, is not yet the case. There are many elaborations of the empirical rules which are the achievement of the past few decades, but few memorable generalizations.

Much the same is true of DNA itself. At least three conformations of the double helix (A, B and Z) are recognized, but there is only the foggiest understanding of how the environment determines the structure of A and/or B, or why particular nucleotide sequences should predispose towards Z instead. Like the protein problem, of course, these are complicated questions, not least because of the certainty that, in molecules in which intramolecular hydrogen bonding is supposed to be a crucial determinant of structure, interactions with extrinsic hydrogen-bonded water molecules are at least as important, but are not yet described.

For lack of the means to tackle these complications, people tackle instead the problems of anhydrous proteins and anhydrous DNA. It is like the drunk who, asked why he searched for his lost latch-key beneath a street-lamp, explained that "It's too dark where I dropped it".

Reflection shows that much the same is

happening elsewhere within the ambit of the programme to make molecular biology the means of explaining life. There is, for example, energy transduction. The molecular structure of muscle fibres is exquisitely described, as is the role of calcium ions, but who can give a simple account of the thermodynamics of the muscle cell? Is molecular biology in that phase corresponding to the introductory phases of infantry training when people are required to learn the names of the pieces of a machine-gun, but are not told how they function or what the whole assembly is for?

It is not essential that the reductionist programme should say how life began, but its successes would be more widely appreciated if it could provide a kind of existence theorem. But instead there seems to have been a succession of fashions. In the 1950s, electrical discharges in mixtures of ammonia, methane and water were shown to yield modest amounts of glycine; in the 1960s there were speculations that the polymerization of cyanogen followed by its hydrolysis would yield even more remarkable products; and there is now a general opinion that the first self-replicating molecules must have been RNA.

Fashion also shapes what is understood of the working of the brain. In less than a quarter of a century, we have been offered the promise that neurotransmitters would tell how the brain works and then that small peptides would do the trick. Now, neural networks have come into their own. Of course, the recognition of the role of dopamine in the substantia nigra has helped in the treatment of parkinsonism (and even, perhaps, of schizophrenia), while neural networks are the most illuminating metaphor of the brain so far. But how well do we understand how the brain works?

The most conspicuous gaps in the molecular description of life are the answers to simple questions such as that. Or, perhaps more rudimentary, what provokes a cell to divide?

Is there a danger, in molecular biology, that the accumulation of data will get so far ahead of its assimilation into a conceptual framework that the data will eventually prove an encumbrance? Part of the trouble is that excitement of the chase leaves little time for reflection. And there are grants for producing data, but hardly any for standing back in contemplation.

John Maddox

Behavioural ecology

Lekking in Florence

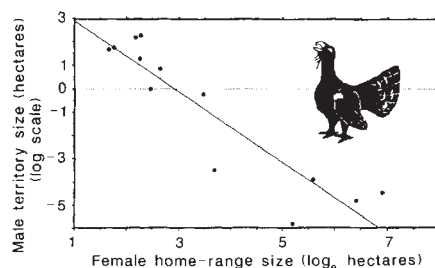
John R. Krebs and Paul H. Harvey

NEW data bearing on the evolution of lekking, the aggregation of reproductively active males at a traditional site, were presented at a recent meeting*. In the breeding season, males gather at the lekking site where they defend small territories (from a few centimetres across in insect leks to a few metres across in lekking species of antelope) and display vigorously to visiting females. Females generally visit the lek at a certain time of day and wander among the males before they mate. The male plays no further part beyond insemination in the female's reproductive effort. Lekking behaviour is taxonomically widespread, occurring in insects, amphibians, birds and mammals, although it is not found in many species within any of these taxa. A striking feature of all leks that have been studied in detail is that a small proportion of the males on a lek obtain most of the matings: in white-bearded manakins (*Manacus manacus trinitatis*) 3 out of 10 males on a lek obtain more than 90 per cent of matings¹, whereas in uganda kob (*Adenota kob thomasi*) 5 out of 22 males obtained more than 80 per cent of matings².

Why does lekking occur in some species but not others? J. Bradbury and S.L. Vehrencamp (University of California, San Diego) with R.M. Gibson (University of California, Los Angeles) investigated the sage grouse (*Centrocercus urophasianus*) of Mono County, California. Bradbury suggested that a major determinant of whether or not a species from the grouse family (Tetraonidae) forms leks is the dispersion of females. If there are places in the environment where females tend to aggregate, for reasons independent of mating, leks will tend to form in these areas (which Bradbury calls hotspots³) because they are the best places to try to mate.

The presence or absence of hotspots is in turn determined by an interaction of two factors: female home-range size and the 'depletion' of available females by males. Consider first home-range size: the chances that the ranges of two or more females will overlap increases as the home-range size of each female increases. These zones of overlap will become hotspots where males aggregate. As home-range size increases further towards the extreme point at which every female range covers the whole area, the extent of overlap zones will increase until the world is one big hotspot and there is no advantage

to a male in focusing his activity in one place. The second factor, depletion, works to decrease the number of hotspots: if a female's home range overlaps with the home ranges of two other females in separate parts of her territory, forming two hotspots, males settling on one hotspot will 'devalue' the second hotspot because one of the two females contributing to it is already being competed for. From a simulation model incorporating



Male territory size decreases with increased female home-range size across 13 species of lekking grouse, a result predicted by Bradbury's hotspot model³ for the evolution of leks.

these two effects, Bradbury predicted that, over the range of values found in the wild for members of the grouse family, males should concentrate into fewer, larger leks as female home-range size increases. As a consequence of crowding into larger leks, Bradbury argues, males should defend smaller territories.

Both these patterns are found in a survey of literature on 13 species of grouse: males occur in fewer, larger leks and have smaller territories in species where female home-range size is larger (see figure). Differences in home-range size are presumably related to differences in some aspect of the ecology of the various species: lekking grouse are open country species and non-lekking species occur in forests⁴. The hotspot hypothesis is not the only one that has been put forward to account for the evolution of leks⁵, but it seems to be a plausible mechanism for the grouse family.

What criteria do females use to distinguish among the males on a lek? There are no correlations between mating success and any measured morphological trait of individual males, such as size, but there is a strong correlation between success and display behaviour (Vehrencamp and Gibson). The male sage grouse has an extraordinary 2-second display called the 'strut', which consists of rapid wing beats, an intense popping sound and a frequency-modulated whistle followed by another popping sound. The strut is associated

with the pumping up of two inflatable orange airsacs which, surrounded by white feathers, look like two large fried eggs on the male's chest. Males vary enormously in their strutting activity: some males do hardly any at all, whereas the real enthusiasts strut at 10-second intervals for 2–3 hours a day over a 2-month period — more than 50,000 struts per season. This effort is not in vain: a multiple regression analysis with mating frequency as the dependent variable shows that successful males put in regular hours at the lek, have a wide frequency-sweep whistle and a long pop-pop interval.

How females might benefit from choosing certain males is a particularly controversial and difficult question to answer using field data. Because the male contributes only sperm to the female, it seems likely that the benefit is genetic: females choose males with a favourable genetic endowment for either survival or mating or both. The finding that the same male sage grouse tend to be chosen by females in successive years is consistent with the notion that there could be genetic variation underlying inter-male variability.

A further line of evidence that the females are choosing males on the basis of a display that indicates male 'quality' comes from measurements of energy expenditure. By using the double-labelled water technique^{6,7}, Vehrencamp showed that 70 per cent of the variation between males in average daily metabolic expenditure (ADMR) is accounted for by the amount of strutting activity and that the most vigorous (and successful in mating) male studied has an ADMR more than 4 times greater than basal metabolic rate (BMR) over a 7-day period. Other studies show⁸ that 4 times BMR is the maximum sustainable work rate for birds and mammals. Hence female choice may have pushed male display behaviour to its physiological limit so that differences in quality between males are amplified. What are the differences in quality that allow different males to display with different levels of vigour? Vehrencamp argues that the answer lies in feeding efficiency: the most vigorous displays travel further to better feeding sites and are thus able to sustain a higher energy expenditure with less weight loss than the less vigorous displays. This final step in the argument remains speculative because of the difficulty in disentangling cause and effect. Nevertheless, this work represents one of the most detailed analyses to date of the relationship between variation in behaviour and mating success in a lek species.

Mating success in the sage grouse is unrelated to position of the male in the lek, but in other species, such as the fallow deer (*Cervus dama*), females seem preferentially to mate in particular territories on the lek rather than with a particular male. The work of M. Apollonio (National

*The Evolution and Ecology of Social Behaviour organized by the Florence Centre for the History and Philosophy of Science, Florence, Italy, 19–24 March 1988.

Institute of Wildlife Biology, Bologna) and colleagues shows that in successive years the same territories in a fallow deer lek were preferred by females, even though the owner may have changed. Two possible explanations of the advantage to females of choosing a particular site are, first, that the site itself is beneficial, for example because it harbours an escape route from predators; and, second, that it is an indirect way of choosing high-quality males by forcing males to compete for favoured sites. Apollonio's data do not exclude the second possibility because it is likely that successive owners of the good sites will be strong competitors.

Lek mating systems are not based on defence by males of resources such as food that are crucial to reproduction, but in other mating systems defence of territories containing resources plays a key role. Dunnocks (*Prunella modularis*) in the Cambridge botanical gardens can form breeding groups that are monogamous pairs, polygynous (one male and two females), polyandrous (one female and two males), or polygynandrous (two females and two males) (N.B. Davies, University of Cambridge). As in Bradbury's account of lekking in grouse, Davies stresses the importance of female home-range size (in turn related to food density) in determining the mating system. When female ranges are small, one male's territory can encompass two female ranges and polygyny results, whereas at the other extreme polyandry occurs when a female's range is large relative to male territories. Monogamy and polygynandry occur at intermediate female range sizes. Superimposed on this ecological background are behavioural conflicts among males and among females to attain the highest possible reproductive output. Males do less well in polyandrous mating groups (because they have to share a mate) and in this circumstance each male constantly battles to evict his rival. Similarly, females in polygynous groups try to evict each other to avoid the cost of sharing a husband's parental care. Davies argues that the observed distribution of mating systems in his dunnocks (and perhaps in many other species) reflects the balance between females pushing towards polyandry and males pushing towards polygyny.

These discussions of mating systems and other aspects of behaviour were all couched in terms of the adaptive advantage to individuals (or their genes) of alternative modes of behaviour. There were more general discussions at the meeting concerning how well adapted one might expect organisms in nature to be. The accepted opinion was that adaptation must be viewed as subject to constraints of, for example, evolutionary history, genetic architecture and development. This is not to say that selection cannot

produce beautiful adaptations. A. Houston (University of Oxford) summarized the point by mentioning Filippo Brunelleschi's exquisite cupola on the Church of Santa Maria del Fiore in Florence. Although the dome appears as a harmonious and integral part of the design of the church, Brunelleschi had to work within the severe constraints of history: his predecessors Arnolfo di Cambio and Francesco Talenti had finished most of the building except for the dome, which appeared to be beyond the technical limits of building at the time. Brunelleschi's miraculous

achievement is, like many of the products of natural selection, an example of optimization within constraints. □

1. Lill, A. Z. *Tierpsychol.* **36**, 1–36 (1974).
2. Floody, O.R. & Arnold, A.P. *Z. Tierpsychol.* **37**, 197–212 (1975).
3. Bradbury, J., Gibson, R.M. & Tsai, I.M. *Anim. Behav.* **34**, 169–179 (1987).
4. Wiley, R.H. *Q. Rev. Biol.* **49**, 201–227 (1978).
5. Bradbury, J.W. & Gibson, R.M. in *Mate Choice* (ed. Bateson, P.P.G.), 109–138 (Cambridge University Press, 1983).
6. Lifson, N. & McClintock, R. *J. theor. Biol.* **12**, 46–74 (1966).
7. Nagy, K.A. *Eco. Monogr.* **57**, 111–128 (1987).
8. Drent, R.H. & Daan, S. *Ardea* **68**, 225–252 (1980).

John R. Krebs and Paul H. Harvey are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Cosmology

Stringing along the galaxies

Edmund Bertschinger

SINCE astronomers announced^{1,2} strong evidence of significant departures on large scales from uniform Hubble expansion of the Universe, cosmologists have been struggling to find an explanation for the formation of universal structure. The recent report³ that the flow of galaxies is dominated by the gravity of a massive 'great attractor' has not made this task any easier. If the attractor is as great as the observers claim, then the popular biased cold-dark-matter (CDM) model for the formation of structure is in serious trouble^{4,5}. Hoffman and Zurek now suggest on page 46 of this issue⁶ that the great attractor is a rare large loop of cosmic string. This exotic possibility has some advantages over alternative explanations.

Inhomogeneities in the Universe, such as the clusters and superclusters of galaxies and correlations of galactic motion, give important signs of its evolution. In particular, Lynden-Bell *et al.*³ observed that superimposed on the uniform Hubble expansion that increases the separation between 400 nearby galaxies, there is a streaming towards a single point — the 'great attractor'. This throws new light on, and introduces new difficulties for, the origin of the structure of the Universe.

Most cosmological theories assume that cosmological structures formed from fluctuations in mass density generated by quantum fluctuations during a period of inflationary expansion in the early Universe. This assumption leads to a unique prediction for the shape of the power spectrum of density fluctuations on large scales, the Harrison-Zeldovich-Peebles scale-invariant spectrum. The amplitude of these fluctuations is normalized by requiring a satisfactory model of galaxy formation and clustering on smaller scales. This procedure leads to an impressively successful theory of galaxy formation from cold, dark matter⁷. Unfortunately, the scale-invariant spectrum seems to have

too little power on large scales to account for the great attractor. Modified power spectra could yield satisfactory agreement with large-scale flows⁸, but there is no compelling mechanism for breaking the scale-invariance in the desired way.

An alternative model for the formation of structure is based on cosmic strings, thin tubes of false vacuum left over from a phase transition in the early Universe. The formation of cosmic strings is a speculative but plausible outcome of 'grand unified theories' of particle physics⁹. In this scheme cosmic strings — mostly in the form of closed loops of all sizes up to the present horizon — accrete the surrounding matter to form galaxies, galaxy clusters and superclusters. This model has not been examined in nearly as great detail as the inflationary CDM model. But it deserves further consideration because, Hoffman and Zurek propose⁶, the great attractor could be explained rather naturally as a cosmic-string loop with a mass of about 10^{13} solar masses ($M_{\odot}$) and a radius of about 50 kiloparsecs (kpc).

Previous calculations^{10,11} of large-scale motions in cosmic-string models considered typical peculiar velocities (deviations from Hubble expansion) averaged over a large volume. The resulting flows (of about 150 km s^{-1} on a scale of $5,000 \text{ km s}^{-1} H_0^{-1}$, where H_0 is Hubble's constant) are comparable with those predicted in the standard CDM model and fall far short of those observed² (about 600 km s^{-1}). Much larger velocities are extremely rare in the CDM model and other models based on random-phase density fluctuations (being in the tail of a gaussian probability distribution) but are not as improbable in the cosmic-string model. Hoffman and Zurek show in this issue⁶ that there is about a 1 per cent probability of finding a sufficiently large loop as close as the great attractor — a much more likely event (assuming that cosmic strings of the proper linear mass-density exist) than

finding a sufficiently large density peak in the biased CDM model.

Hoffman and Zurek also argue that the velocity field generated by a cosmic-string loop (infall velocity $v \propto r^{-1}$ steepening to r^{-2} at large distances r from the loop) agrees better with the great attractor flow field³ than that predicted by the CDM model ($v \propto r^{-3}$ at large distances). This argument is weakened by the revised great-attractor model of Faber and Burstein¹² which has $v \propto r^{-1.7}$ outside a core with a radius of about $1,400 \text{ km s}^{-1} H_0^{-1}$. But Hoffman and Zurek show that the flow field is modified by the expected motion of the loop. From their results it seems likely that they can construct a satisfactory, detailed model of the great attractor.

What are the consequences of a cosmic-string loop as great attractor? The loop should still be present — not having yet decayed entirely to gravitational radiation — subtending an angle of about 2 arc min. It should lie within a few degrees of the great attractor, but need not coincide with the centre of the galaxy concentration because of its own peculiar velocity. The loop should be detectable by gravitational lensing — a cosmic string produces double images of distant stars with separations of a few arc s — but it could be obscured by dust in the plane of our Galaxy. A moving loop would not necessarily create a dense galaxy cluster detectable by X-ray emission from intracluster gas, but a $10^{13}\text{-}M_\odot$ object smaller than 100 kpc should form a dense bound object of some kind. The many other string loops that should exist according to this model would produce gravitational radiation and microwave background fluctuations which should be detectable within a few years.

The cosmic-string model is considered by most cosmologists, myself included, to be a long shot. But as other models for the formation of large-scale structure are pressed increasingly by observations like that of the great attractor, prudence would seem to dictate that we consider alternatives to the usual schemes of inflation. □

1. Collins, C.A., Joseph, R.D. & Robertson, N.A. *Nature* **320**, 506–508 (1986).
2. Dressler, A. *et al. Astrophys. J.* **313**, L37–L42 (1987).
3. Lynden-Bell, D. *et al. Astrophys. J.* **326**, 19–49 (1988).
4. Vittorio, N., Juszkiewicz, R. & Davis, M. *Nature* **323**, 132–133 (1986).
5. Vittorio, N. & Turner, M.S. *Astrophys. J.* **316**, 475–482 (1987).
6. Hoffman, Y. & Zurek, W.H. *Nature* **333**, 46–49 (1988).
7. Davis, M., Efstathiou, G., Frenk, C.S. & White, S.D.M. *Astrophys. J.* **292**, 371–394 (1985).
8. Bardeen, J.M., Bond, R.J. & Efstathiou, G. *Astrophys. J.* **321**, 28–35 (1987).
9. Vilenkin, A. *Phys. Rep.* **121**, 264–315 (1985).
10. Shellard, E.P.S., Brandenberger, R.H., Kaiser, N. & Turok, N. *Nature* **326**, 672–675 (1987).
11. van Dalen, A. & Schramm, D.N. *Astrophys. J.* **326**, 70–76 (1988).
12. Faber, S.M. & Burstein, D. *Proc. Pont. Acad. Sci. Study Week 27* (in the press).

Edmund Bertschinger is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Oxide superconductors

Magnetic effects explored

V.J. Emery

EVIDENCE for the close connection between magnetism and high-temperature superconductivity is steadily mounting. Two new experiments reported by Lyons *et al.*¹ and Brewer *et al.*² bolster the belief that the copper oxide planes in doped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ behave in essentially the same way as those in doped La_2CuO_4 : magnetic order gives way to short-range magnetic correlations and superconductivity as the dopant or oxygen concentration is increased. There are indications that the magnetic moments of the copper ions persist into the superconducting phase.

Many have suggested that coupling of the conduction electrons to the magnetic moments in the new superconductors could cause the pairing essential to superconductivity. Experiments carried out over last year revealed striking magnetic behaviour in the copper oxide planes of La_2CuO_4 (see my discussion in News and Views at that time³). Lyons *et al.*¹, using inelastic light scattering to excite pairs of spin waves (magnetic oscillations), find evidence of antiferromagnetism in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. Muon-spin rotation experiments performed by Brewer *et al.*^{2,4} show that there is magnetic ordering in this material for x less than about 0.4. Several groups (ref. 5; W.-H. Li *et al.*; J. Rossalt Mignod *et al.*; and J. Petitgrand and G. Collin, preprints) have now used neutron scattering to determine the antiferromagnetic structure of this material.

The essential features are very similar for the parent materials of both the 40-K (La_2CuO_4) and 90-K ($\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$) superconductors: large intrinsic ordered moments of 0.5–0.6 Bohr magnetons (μ_B) and quite high exchange interactions of 120–140 meV. To be sure, the magnetic structures of the ordered phases are rather different, and $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ has a higher maximum magnetic-ordering (or 'Néel') temperature T_N . But these properties are largely a reflection of extrinsic factors — the crystal structure and the magnetic coupling between the planes.

Up to now, there is no evidence that antiferromagnetic order and superconductivity coexist in the same sample. In $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, the two phases are separated by a region, extending from about 2 to 5 per cent doping, which has many of the characteristics of a spin glass (refs 6,7; D.R. Harshman *et al.*, preprint). On the other hand, near the borderline oxygen concentration ($x=0.4$), $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ is usually found to be either antiferromagnetic or superconducting. But the behaviour probably depends upon the method of preparation, and particularly

the way in which the oxygen content was established: one sample used by Brewer *et al.*² was superconducting below the transition temperature $T_c=25 \text{ K}$ and then ordered magnetically below $T_N=10 \text{ K}$, another had $T_c=33 \text{ K}$ and $T_N=5 \text{ K}$. There is some inconclusive evidence that antiferromagnetism destroyed the superconductivity. Perhaps the greatest problem is to allay the suspicion that the borderline samples are structurally heterogeneous, which would make it extremely difficult to extract the true behaviour of a uniform system.

The central question is the extent to which extended magnetic correlations, if not order, persist as the doping or oxygen content is changed to make the sample superconducting. There is some evidence that they do, although, once again, sceptics will question the homogeneity of the material. Light scattering¹ in two samples of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, one with $x \approx 0.6$ and $T_c=60 \text{ K}$, the other having $x \approx 0.9$ and $T_c=88 \text{ K}$, shows broad peaks similar to those produced by the excitation of pairs of spin waves in a diluted antiferromagnetic insulator. Moreover, neutron-scattering experiments⁸ on $\text{La}_2\text{Cu}_{0.95}\text{Li}_{0.05}\text{O}_4$ and $\text{La}_{1.97}\text{Sr}_{0.03}\text{Cu}_{0.95}\text{Li}_{0.05}\text{O}_4$ which are not superconducting, show that there are indeed magnetic correlations but they extend over much shorter distances than in the undoped material. The most important observation, however, is that the integrated intensity of the peaks is unaffected by doping⁹, which suggests that the charge carriers interact strongly with the local magnetic moments and modify their behaviour, but do not destroy them.

What are the implications of all of this for the behaviour of the copper oxide planes? The ordered moment of 0.5–0.6 μ_B for the undoped material is very close to that expected for localized spins undergoing little more than quantum spin fluctuations. Any substantial translational motion would bring about a further reduction of the moment below its observed value. It is also difficult to escape the conclusion that the order is an intrinsic property of the ground state of the individual planes: the weak interplane coupling is crucial at finite temperatures but cannot, of itself, give rise to such a large moment⁹. Such an ordered ground state is consistent with the temperature dependence of the range of the magnetic order⁹ and should lead to characteristic neutron scattering above the actual magnetic-ordering temperature T_N , where the three-dimensional coupling is less important. Indeed, seemingly elastic

scattering has been observed⁸, although it is much weaker than in magnetically anisotropic higher-spin systems, where it dominates.

On the other hand, much remains to be done to elucidate the evolution of the magnetic state as charge carriers are introduced by doping or varying the oxygen content. It is important to confirm the current trends and accumulate the evidence to show that the local moments really do persist into the superconducting phase. One thing is clear; the magnetism cannot be ignored. It surely is the reason why the charge carrier density is so low,

and even quenching of the moments would have specific implications for the nature of the superconducting state. □

1. Lyons, K.B. *et al.* *Phys. Rev. Lett.* **60**, 732–735 (1988).
2. Brewer, J.H. *et al.* *Phys. Rev. Lett.* **60**, 1073–1076 (1988).
3. Emery, V.J. *Nature* **328**, 756–757 (1987).
4. Nishida, N. *et al.* *Jap. J. appl. Phys.* **26**, L1856–L1858 (1987).
5. Tranquada, J.M. *et al.* *Phys. Rev. Lett.* **60**, 156–159 (1988).
6. Tokumoto, M., Nishihara, Y., Oka, K. & Unoki, H. *Nature* **330**, 48–49 (1987).
7. Aharony, A. *et al.* *Phys. Rev. Lett.* **60**, 1330–1333 (1988).
8. Endoh, Y. *et al.* *Phys. Rev.* (in the press).
9. Chakravarty, S., Halperin, B.I. & Nelson, D.R. *Phys. Rev. Lett.* **60**, 1057–1060 (1988).

V.J. Emery is at Brookhaven National Laboratory, New York 11973, USA.

Surface science

Methods of time-resolved study

David A. King

REAL-TIME surface spectroscopic studies would be invaluable for the investigation of the many dynamic phenomena that occur on surfaces. H.E. Elsayed-Ali and G.A. Mourou now report (*Appl. Phys. Lett.* **52**, 103–104; 1988) a reflection high-energy electron-diffraction (RHEED) pattern obtained from the cleaved surface of a sodium chloride crystal using a single 200-picosecond (ps) electron pulse (Fig. 1). Although the inherent repeat time in their experiment is limited by the detectors they use — a phosphor screen with a persistence time of 0.15 ms and a microchannel plate with an even longer dead time — this type of development heralds a new generation of tools for surface science.

The relatively slow development of time-resolved techniques in surface science has a simple explanation. There are only about 10^{15} atoms or molecules (10^{-9} moles) cm^{-2} at a well-defined surface available to the experimentalist. Although a wide range of particle-based techniques have now been developed using combinations of electron, photons, ions and atoms

as incident and exit particles, intense particle sources are often needed simply to obtain sufficient sensitivity to detect surface species. Some techniques have very limited surface sensitivity except under restrictive conditions: for example, X-ray diffraction is useful only with close to grazing incidence. But because electrons with energies of 50–3,000 eV have mean free paths of only a few atomic diameters in solids, techniques based on detecting exit electrons are sensitive only to those electrons that are generated near the surface of the solid and have a well-defined energy.

For example, low-energy electron diffraction (LEED) operating at 50–300 eV, is inherently surface-sensitive, and is a strong candidate for time-resolved studies. This technique was applied a few years ago (Becker, R.S., Higashi, G.S. & Golovchenko, J.A. *Phys. Rev. Lett.* **52**, 307–310; 1984) to the study of the dynamics of laser annealing of germanium. But at these low electron-beam energies, the current is space-charge limited (Coulomb repulsion between the electrons causes the beam to diverge), and it has been estimated that only one electron is reflected from the surface for each nanosecond shot.

For the study of adsorbates on surfaces, vibrational spectroscopy offers exciting possibilities, given the wealth of information that it provides. One such method however, high-resolution electron-energy-loss spectroscopy (HREELS), is usually operated at lower electron energies than LEED, and therefore runs into the same limitations on beam current. The best time

resolution achieved with this technique to date (Dubois, L.H., Ellis, T.H., Zegarski, B.R. & Kevan, S.D. *Surf. Sci.* **172**, 385–397; 1986) is 5 ms at constant frequency. Predictions of spectral scans with millisecond time resolution in HREELS have not yet been realized. Reflection-absorption infrared spectroscopy (RAIRS) offers greater potential, although the inherent noisiness of tunable infrared lasers limits its present use. Absorption bands are weak and easily lost in the noise. The best time resolution to date with RAIRS (30 ms) has been obtained with a Fourier-transform instrument, with a traditional black-body infrared source (Chesters, M. *J. elec. Spec. Related Phen.* **38**, 123–140; 1986).

In the new system described by Elsayed-Ali and Mourou, the 25-keV electron source is a photocathode, composed of a gold film deposited onto a sapphire window, activated by pulsed ultraviolet light from a YAG laser (Fig. 2). For studies of dynamic events, this source provides a good pulse width (about 150 ps) and beam current (5×10^5 electrons per pulse), but the maximum repetition rate is limited by the detectors to 1.5 kHz. Further develop-

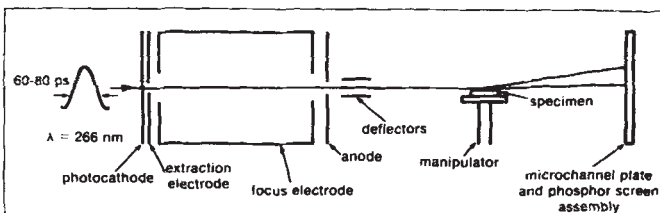


Fig. 2 Schematic diagram of the system used by Elsayed-Ali and Mourou for time-resolved studies of crystal surfaces. A picosecond laser pulse excites a pulse of electrons from the photocathode, which is accelerated, focused and scattered off the sample surface.

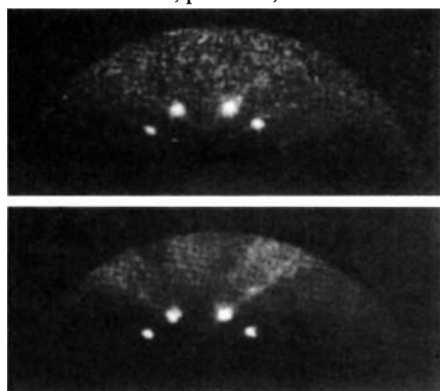


Fig. 1 Reflection high-energy electron diffraction patterns from a cleaved surface of a sodium chloride crystal recorded with a microchannel plate. a, Using a single 200-ps electron pulse; b, averaged over 50 pulses. (From Elsayed-Ali and Mourou.)

ments in rapid detection are necessary for repetitive studies of dynamic events.

The immediate need for real-time surface techniques arises from the important development of pulsed laser processing of metals and semiconductors, and the need to monitor the transformations in these materials. But the whole field of dynamics at surfaces, including surface melting and roughening, surface diffusion, local density and structural fluctuations, the formation of metastable adsorbate states and short-lived catalytic intermediates, is ripe for detailed study once these techniques become available. In studies of adsorbates, pulsed laser time-of-flight mass spectrometry and secondary ion mass spectrometry hold great promise, particularly in conjunction with pulsed molecular beam sources. For alloy systems, ion-scattering spectroscopies could be developed with the availability of intense, pulsed ion sources. For structural studies, RHEED is currently the leading contender. □

David A. King is in the Department of Physical Chemistry, Cambridge University, Lensfield Road, Cambridge CB2 1EP, UK.

Vision

Perspectives on movement

Brian Rogers

AN ESSENTIAL requirement of any biological or machine-vision system is the ability to detect the motion of images across the photosensitive receptors or transducers in the eye or camera. Our understanding of motion perception has benefited significantly from the interplay between the results of psychophysical studies using human observers on the one hand, and the electrophysiological and anatomical studies of the visual cortex of cats and monkeys on the other. A third approach can now be added — the computational — which seeks to provide a mathematical framework for understanding motion perception. The complementary nature of these three approaches is evident in the work of Yuille and Grzywacz, reported on page 71 of this issue¹, which outlines a computational theory of motion perception that can account for various psychophysically observed 'errors' of human motion perception. The importance of mathematics in the description and analysis of image motion was also an important theme at a recent conference*.

The fundamental problem of motion perception is that any individual receptor or transducer can signal information only about the amount of light reaching it at any moment. Motion perception, however, is necessarily a spatio-temporal process which requires the detection of changes of light falling on two or more

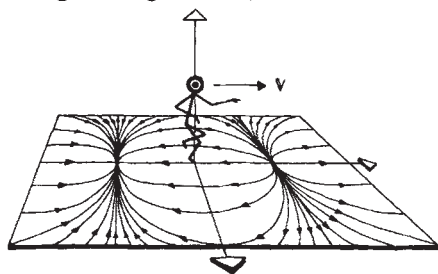


Fig. 1 The flow patterns generated when an observer moves through an environment parallel to the plane. (Courtesy of J. Koenderink, from ref. 11.)

spatially separated regions of the receptor surface over time. Using data derived from the optokinetic responses of the fly, Reichardt and Hassenstein about 30 years ago proposed² an autocorrelation model of motion perception in which the spatially and temporally filtered outputs of pairs of detectors were multiplied to provide a signal of image motion. Sequence-detecting models of motion perception^{3,4} and the recent gradient models^{4,5}, however, all suffer from the limitation that they can detect signal only

in a direction orthogonal to the orientation of a luminance contour in the image. This has been called the 'aperture' problem. W. Reichardt proposed that a network of simple correlating mechanisms would be immune to this problem.

In their paper in this issue¹, Yuille and Grzywacz suggest that the problem could also be solved by summing, and thereby smoothing, the motion signals derived from a set of simple motion detectors over

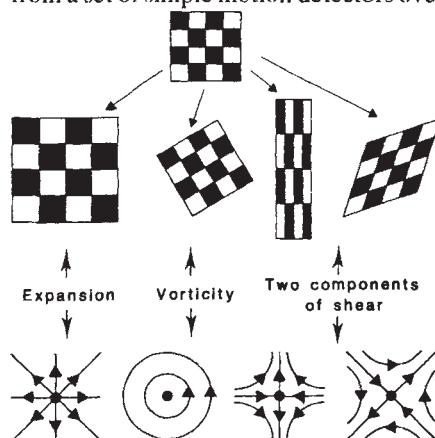


Fig. 2 The four differential invariants into which any flow field can be decomposed. (Courtesy of J. Koenderink, from ref. 11.)

a limited spatial region of the moving image. This solution is different from that proposed by Hildreth⁶ in which smoothing takes place along contours.

The psychophysical or behavioural evidence to support this kind of smoothing operation in the human visual system was first noted by the Gestalt psychologists 50 years ago, who showed that we perceive the common direction of moving elements even when the elements have some degree of additional random motion — as with smoke rising from a cigarette or chimney.

Although the proposed computational model accounts for these 'common-fate' percepts, and also the phenomena of motion cooperativity and motion capture, several important characteristics of perception are not accounted for. First, we do not find it difficult to detect the sharp boundaries between moving and non-moving regions, which are effectively blurred out by Yuille and Grzywacz's algorithm (their Fig. 1b on page 72) and it is known that the visual system uses motion to segment different parts of the scene⁷. Second, although their model correctly identifies the direction of bias in a field of random motion, it does so at the expense of eliminating the random motions of the elements, which are easily visible in such a display. Clearly, their algorithm must be extended to limit the

amount of smoothing that can occur across discontinuities, and there have to be parallel mechanisms which preserve the perturbations of any motion field if it is to account for the subtleties of human motion perception.

Patterns of motion in images are not only useful for providing information about the movements of objects in the surrounding world, they also provide information about the perceiver's own movements. Indeed, most of the motion present in our retinal images results from our own eye, head and body movements, rather than from the movements of objects in the world. At the meeting, J. Koenderink (University of Utrecht) proposed that image flow (to distinguish it from the optic flow measured against some independent coordinate framework) might be detected using Reichardt-type correlator mechanisms over large areas of the retina. The derived flow patterns could then provide information about our own rotations and translations with respect to the world. For example, a particular flow pattern could tell us about our direction of heading within a few degrees⁸ and the time before reaching a particular surface⁹ (see figures). Recent behavioural work (H. Dahmen and W. Junger, University of Tübingen) shows that waterstriders, insects which live on still or slowly moving water surfaces, are able to use the characteristics of image flow to distinguish between passive rotations or translations to which they are subjected. In addition, there is now physiological evidence (J. Rauschecker and R. Kreiter, Max Planck Institute, Tübingen) that there are neurons in the lateral suprasylvian cortex of the cat which prefer centrifugal motion away from the area centralis and also respond to whole-field expansion or contraction patterns of the sort generated when the animal moves through the visual world. In addition, anatomical work using retrograde tracers shows that there are long-range intercortical connections which could support the global analysis of image flow.

Although it is often assumed that the fine three-dimensional structure of surfaces and objects could also be extracted from the same global velocity flow fields (structure from motion), Koenderink suggests that the information could be obtained more accurately by detecting the local distortions or deformations within

1. Yuille, A.L. & Grzywacz, N.M. *Nature* **333**, 71–74 (1988).
2. Reichardt, W. *Z. Naturforschung* **12b**, 447–457 (1957).
3. Barlow, H.B. & Levick, W.R. *J. Physiol., Lond.* **178**, 477–504 (1965).
4. Marr, D. & Ullman, S. *Proc. R. Soc. B* **211**, 151 (1981).
5. Adelson, E.H. & Bergen, J.R. *J. Opt. Soc. Am.* **2**, 284 (1985).
6. Hildreth, E.C. *The Measurement of Visual Motion* (MIT, Cambridge, Massachusetts, 1987).
7. Cutting, J. *Perception with an Eye for Motion* (MIT, Cambridge, Massachusetts, 1986).
8. Braddick, O.J. *Vis. Res.* **14**, 519–527 (1974).
9. Lee, D.N. *Phil. Trans. R. Soc. B* **290**, 169–179 (1980).
10. Koenderink, J.J. *Vis. Res.* **26**, 161–180 (1986).
11. Ingle, J. et al. (eds) *Brain Mechanisms and Spatial Vision* (Nijhoff, Dordrecht, 1985).

*European Brain and Behaviour Society Workshop on Visual Processing of Form and Motion Tübingen, 16–19 March 1988.

the flow field. In particular, the amount of local shear or deformation contains information about the slant and tilt of surface patches. Deformation could be extracted or computed by detecting the changes in orientation of local line segments, without the need to detect motion *per se*. The oriented receptive fields found in the primary visual cortex of cat and monkey could well be involved in computing surface slant and structure from image flow in this way¹⁰.

Higher-order shape characteristics, such as the curvature of surfaces, could also be detected using the local deformations in the flow fields generated when an observer moves with respect to a three-dimensional surface. The second spatial derivative of the flow field (the change of motion gradient across space), created by a curved three-dimensional surface, has the interesting characteristic that it does

not change with the viewing distance to the surface (my own work). To extract the motion curvature in the flow field, the visual system would not necessarily have to differentiate the velocity field twice with respect to space (which sounds a rather unlikely biological operation). Instead, motion curvature could be computed by detecting the changes in the curvature of local line segments in the image, which has the advantage of being both more plausible from a biological point of view and more accurate than computations based on velocity vectors. At present, these ideas are speculative, but the potential for investigating these and other problems using psychophysical, physiological and computational techniques is a reason for optimism. □

Brian Rogers is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

Oceanography

Effects of microbe activity

Erwin Suess

ORGANIC matter exists in many forms in the oceans — as detritus and microbes, which can be suspended or sinking, and also as dissolved organic matter. The way these pools interact is influenced by, and influences, the distribution of nutrients and oxygen dissolved in the water column. One point of current interest is the fate of large sinking phytodetritus particles and their interaction with microorganisms in the ocean's interior and at the sea floor. In two papers recently published in *Nature*, Karl *et al.*¹ and Cho and Azam² show that microbial activity on sinking phytodetritus decreases with depth and is minimal below 2,000 m in the ocean. On the other hand, on page 67 of this issue³, Lochte and Turley show that phytodetritus, once it has reached the sea bed, hosts vigorous microbial activity.

The resolution of these contrasting results bears on several important issues in marine science. One long-standing problem is that the downward flux of sinking organic particles decreases with increasing depth. What is the fate of the lost particles? To what extent are they broken up into smaller suspended particles or else remineralized by microbes, releasing carbon dioxide and soluble nutrients, consuming oxygen in the process? Another problem concerns the microbial activity at and near the sea floor. This process can affect the local oxygen distribution and that mixed into the water column above, having consequences for palaeo-oceanography and ocean chemistry.

Lochte and Turley³ demonstrate in this issue that large, fresh particles of phytodetritus recovered from the sea floor at

4,500 m depth in the north-east Atlantic are populated by heterotrophic bacteria, cyanobacteria and microflagellates (Fig. 1). The nutritious host particles must have arrived at the deep-sea floor only shortly before they were sampled because the cells of cyanobacteria, photosynthetically supported picoplankton that inhabit surface waters in summer, remain intact. The microbial population on the recovered detritus flourished when incubated under simulated deep-sea and shallow-water



Fig. 1 Cyanobacteria (red) and non-pigmented bacteria recovered from the sea bed. (Photo courtesy of K. Lochte and C.M. Turley.)

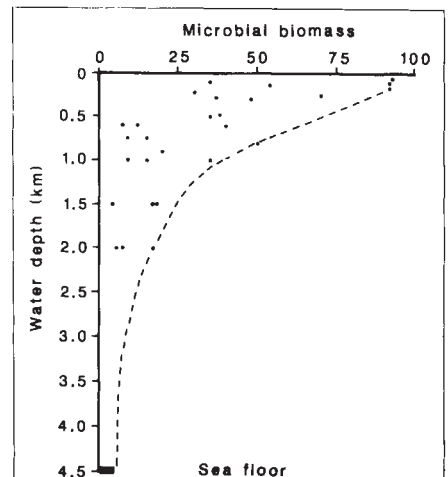


Fig. 2 Estimates of microbial biomass of fast-sinking, organic particles in the ocean^{1,2} and at the sea floor³ as weight per cent of total detritus.

conditions. Its respiratory activity and the rates at which it converted carbon into biomass affected the chemistry and mass of the entire suspended-particle pool. The rate of carbon respiration during the experiment simulating deep-sea conditions indicates that phytodetritus from the sea surface would be consumed within two months of settling on the sea floor.

Microbial populations attached to sinking particles in the water column and estimates of their biomass have been reported before. Biomass estimates from the north-east Pacific, reported by Karl *et al.*¹ and Cho and Azam², extrapolated down the oceanic water column, give a distribution of microbial biomass decreasing with depth to reach the sea-floor value found by Lochte and Turley³ (Fig. 2). This distribution is at odds with the rapid rates of decomposition observed by Lochte and Turley. Karl *et al.* trapped sinking organic particles in the top 2,000 m of the ocean. The microbial population in these consistently died during *in situ* incubation experiments, despite the nutritious environment provided by the host particles.

How then is it possible that the microbial population attached to particles recovered from the sea floor is so unusually active? The incubation experiments provide an interesting possible explanation. The activity of the cyanobacteria initially attached to the particles at the sea surface lasted only a few days, whereas the high activity of heterotrophic bacteria was sustained for the duration of the experiment. Could it be only after the phytodetritus reaches the sea floor that the specialized barophilic heterotrophs populate the particles? These heterotrophs would be better able to use the nutritious detritus than the attached cyanobacteria. Particles recovered from the water column by trapping¹, of course, could not have been populated by sea-floor microbes. Clearly, the mechanism of decomposition at the sea floor is not merely an extension of that

in the water column.

The rapid response of the deep-sea microbes has consequences for palaeo-oceanography. At the same north-east Atlantic site where the microbe-containing particles were collected, certain species of rare benthic foraminifera populated the freshly arrived phytodetritus in great abundance⁴. These species are specially adapted consumers feeding opportunistically as others remain unaffected by the organic influx. Hence, the abundance pattern of benthic foraminifera would be a direct reflection of the organic influx. Also, their habitat would become enriched in metabolic carbon dioxide from respiratory consumption. These conditions could be recorded in the species' assemblages⁵ and the stable carbon isotopes of their calcareous shells⁶.

The deep-sea floor along productive continental margins has been identified by respiration data⁷ as the main site of oxygen consumption in response to organic influx. Previous models of the oxygen distribution, based largely on fast-sinking particle fluxes, favoured the shallow water column as the site of high consumption. The new concept implies that bottom water, after having its oxygen extracted by the adjacent sea floor and being advected laterally, could control the oxygen and nutrient chemistry of the ocean's interior. These studies⁴⁻⁷ show that the rapid turnover of phytodetritus at the sea floor is the key process.

The low microbial activity observed by Karl *et al.*¹ also throws light on the reduction of the flux of sinking organic particles. These particles are poor hosts to attached microbes, which therefore cannot be responsible for the lost flux, as was previously thought. Instead the sinking particles probably disaggregate before free-living microbes turn the suspended fragments into soluble nutrients.

This notion agrees with Cho and Azam's observation² that most of the suspended organic matter in the upper ocean is bacterial. (Specifically, they find that 90 per cent of the surface area of suspended particles is on microbes.) Additional evidence of free-living microbes in the water column comes from Wakeham⁸ who shows that suspended particles are associated with fresh, unaltered steroids typically found in living organisms, whereas sinking particles are associated with steroids altered by extracellular degradation. □

1. Karl, D.M. *et al.* *Nature* **332**, 438–441 (1988).
2. Cho, B.C. & Azam, F. *Nature* **332**, 441–443 (1988).
3. Lochte, K. & Turley, T.M. *Nature* **333**, 67–69 (1988).
4. Gooday, A.J. *Nature* **332**, 70–73 (1988).
5. Lutze, G.F. *et al.* *METEOR-Forschungsergebnisse* **C40**, 163–180 (1986).
6. Zahne, R., *et al.* *Paleoceanography* **1**, 27–42 (1988).
7. Jahnke, R.A. & Jackson, G.A. *Nature* **329**, 621 (1987).
8. Wakeman, S.G. *Geochim. cosmochim. Acta* **51**, 3051–3069 (1987).

Erwin Suess is in the College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA.

Transposition immunity

Action at a distance

Neville Symonds

THE biggest revolution in our thinking about genetics in recent years has been the belated acceptance of the idea first proposed by Barbara McClintock in the 1940s, that there are small genetic elements which can move from one location to another within the genomic DNA present in cells. These mobile elements, now called transposons, have been identified in bacteria, fungi, insects, plants and animals, and shown to be responsible for effects ranging from the spread of antibiotic resis-

antibiotics, spread through bacterial populations and are able to establish themselves in new bacterial strains.

There is now general agreement on some aspects of the phenomenon of transposition immunity^{3,6}. It has become clear that it is not caused by the inability of the plasmid to carry more than one copy of a particular transposon as plasmids containing more than one copy can easily be made. The immunity must therefore result from a block in the successful completion of the transposition cycle. It is also clear that immunity is not dependent on a complete transposon being present within the target. The essential factor is that there is a binding site for transposase. Further, immunity of several orders of magnitude can be obtained with target molecules of up to 100 kb in length. Just a few molecules of transposase located at a single site in a molecule therefore determine the immunity.

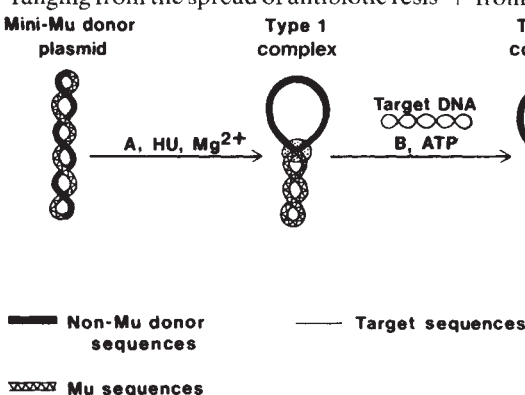


Fig. 1 Steps in the strand-transfer reaction.

tance in bacteria to the onset of cancer in animals. An unexpected observation noted over 10 years ago with bacterial transposons was that insertion of a bacterial transposon into a plasmid (a small circular DNA molecule) within a cell was severely inhibited if the plasmid already contained a copy of the same transposon¹. This phenomenon, called transposition immunity, was first reported for the transposon Tn3, and has since been found in other members of that family, in Tn7 and in phage Mu (refs 2–4). An intriguing feature of transposition immunity is that a single copy of the transposon located at a site within the plasmid can inhibit the insertion of another copy at sites up to 50 kilobases (kb) away⁵. A paper by Adzuma and Mizuuchi⁶ in the 22 April issue of *Cell* gives the first clear account of the biochemical basis underlying an example of transposition immunity, and in so doing opens up ways of investigating this 'action at a distance' aspect of the phenomenon.

Transposons become inserted into new locations in genomes by a special type of recombination process in which specific enzymes (transposases) recognize and bind to unique sequences present at each end of the transposons. Plasmids occur naturally in bacteria and are transferred from one cell to another during bacterial conjugation. They provide the vehicles by which transposons, many of which carry determinants conferring resistance to

Before I describe the new work of Adzuma and Mizuuchi, I shall review briefly what is now known about the molecular nature of transposition in the bacteriophage Mu, as deduced from *in vitro* experiments⁷. The key step in this transposition is one of strand-transfer, leading to the formation of a branched DNA molecule in which the 3' termini of Mu DNA are joined to the 5' termini of a 5-base-pair staggered cut in the target DNA (ref. 7). A defined *in vitro* system that efficiently supports this reaction is now available. It consists of Mu A protein (the transposase), Mu B protein (a non-specific DNA-binding protein which is also an ATPase), *Escherichia coli* HU protein (a histone-like protein), ATP and Mg²⁺. Into this milieu are introduced small circular donor and target DNA molecules, the donor containing a mini-Mu derivative of phage Mu^{7,8}.

Two steps have now been identified in the strand-transfer reaction^{9,10}. In one, the Mu A protein binds to the Mu end sequences and with the HU protein introduces specific nicks at the termini of the Mu DNA, so leading to the formation of a stable type 1 complex in which the Mu ends are held together by protein. In the second, this type 1 complex interacts with target DNA in the presence of Mu B protein and ATP to complete the strand-transfer reaction and form the branched type 2 complex.

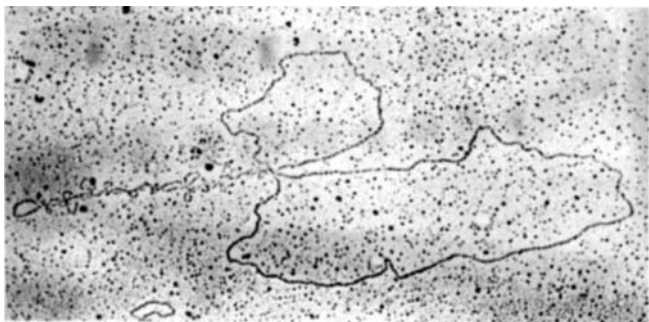


Fig. 2 Electron micrograph of a type 2 complex, magnification $\times 3,520$. Donor molecule plasmid psU1 (37 kb Mu + 14 kb pML2); target molecule plasmid R388 (32 kb). This sample is from induced plasmid-containing maxi-cells (cells in which chromosomal DNA has been degraded by nuclease attack after prior irradiation). (Photo, H. Janzen, University of Sussex).

Type 1 and type 2 complexes appear in the electron microscope as structures similar to those in Fig. 1, in which distinct topological domains can be recognized. Some caution has to be applied, however, in extending these results to the *in vivo* behaviour of phage Mu, which is 37 kb long, as the mini-Mu derivatives used in the *in vitro* system are less than 4 kb long and the donor and target molecules less than 10 kb. Some confidence that the extrapolation can be made comes from the electron micrograph in Fig. 2, which shows a type 2 complex isolated from induced cells which contain as donor a 51-kb plasmid including the whole of phage Mu, and as target a 32-kb plasmid.

Adzuma and Mizuuchi now show⁶ that in the *in vitro* system transposition immunity (which is recognized by the inability of molecules carrying Mu end-sequences to act as efficient targets in the strand-transfer reaction) is correlated with an inability of the immune targets to bind Mu B protein stably. In the initial stages of the usual *in vitro* reaction, stable binding of Mu B protein associated with ATP occurs on the target molecules. But the presence of A protein bound to the target induces dissociation of Mu B from DNA in a process associated with ATP hydrolysis. Target immunity therefore comes about in this system from a shortage of the bound Mu B protein which is necessary to facilitate the second stage in the strand-transfer reaction.

It remains to be seen whether this mechanism of transposition immunity is relevant to other systems. Mu is unique as a transposon in that the specific end-binding protein and the non-specific target-binding protein are distinct. Other transposons have only a single transposase molecule which must combine both these roles, and although it is well known that these transposases have comparable molecular masses to that of Mu A and Mu B combined, and could easily contain multiple domains, it has not yet been shown that any possess ATP-binding sites.

Transposition immunity could also be mediated by other mechanisms. One

suggestion is that the three-dimensional geometry of the site involved in the type 1 to type 2 transition is critical, and that presence of an additional transposase molecule on the target could influence this structure¹¹. As yet, however, neither the ATPase nor the tentative topological explanations provides any clues about the most intriguing aspect of transposition im-

munity — how it can act over such large distances. Adzuma and Mizuuchi are now addressing this aspect of the problem in the *in vitro* system, testing the possibilities that the A and B proteins interact with each other by looping out of intervening DNA, or by continuous movement of B protein along the DNA molecules.

Another question about transposition immunity is what is it for? The simple evolutionary answer is that if there is one copy of a transposon in a genome, there is not much point in adding another. This is probably too extreme a point of view, as certainly with the smaller transposons there are numerous reports of genomes containing more than one copy. However, even in these cases the number is limited, which makes sense as transposition is inevitably a mutagenic process. It seems plausible therefore that transposition

immunity represents one way of controlling the number of transposition copies present within cells. With phage Mu, which replicates its DNA during the lytic cycle by repeated rounds of transposition, an alternative answer is that transposition immunity prevents the self-destructive effects which could result from Mu insertions occurring within other Mu molecules. Mu uses transposition at three stages in its life cycle. After phage infection, the Mu genome integrates into host DNA by transposition. In lysogenic cells Mu is similar to other transposons in its ability to transfer to new locations. And, as mentioned above, Mu replicates its DNA by transposition during phage growth. It is something of a mystery that with all the advantages of this variegated life-style backed up by a sophisticated and efficient transposition system and benefiting also from transposition immunity, Mu nonetheless had become virtually extinct before it was resurrected by Larry Taylor in 1963. □

1. Robinson, M.K., Bennett, P.M., Grinstead, J. & Richmond, M.H. *J. Bact.* **129**, 407 (1977).
2. Hauer, B. & Shapiro, J. *Molec. gen. Genet.* **194** (1984).
3. Arthur, A. *et al. EMBO J.* **5**, 443 (1984).
4. Huang, Chang-Jen *et al. Gene* **41**, 23 (1986).
5. Wallace, L.C., Ward, J.R., Bennett, P.M. & Richmond, M.H. *Molec. gen. Genet.* **184**, 80 (1981).
6. Adzuma, K. & Mizuuchi, K. *Cell* **53**, 257–266 (1988).
7. Craigie, R. & Mizuuchi, K. *Cell* **41**, 867 (1985).
8. Craigie, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7570 (1985).
9. Craigie, R. & Mizuuchi, K. *Cell* **51**, 493 (1987).
10. Surette, M.G. *et al. Cell* **49**, 253 (1987).
11. Sheratt, D. *et al. in Genetic Re-arrangement* (5th John Innes Symp., Croom Helm, London, 1983).

Neville Symonds is Professor of Microbial Genetics at the University of Sussex, Falmer, Brighton BN1 9QG, UK.

Behavioural neurogenetics

Designing heat-shock clocks

Martin Heisenberg

As evidence accumulates that animals may require most of their genes for their brains and hence their behaviour, behavioural genes are becoming a focus of attention. One of the most advanced case studies is that of J.C. Hall, M. Rosbash and collaborators at Brandeis University. It concerns a gene of the fruitfly *Drosophila* named *period* (*per*), which may play a key role in the rhythmicity of behaviour. In their most recent work, described on page 82 of this issue¹, these investigators construct, by *in vitro* mutagenesis (see figure), temperature-sensitive *per* mutants. Their results provide a nice illustration of the impact of genetic tools on behavioural analysis.

An air of excitement has surrounded the *per* gene since its discovery² in the early 1970s. To understand the mechanisms underlying biological clocks, Konopka and Benzer began to look for *Drosophila* mutants with altered circadian

rhythms. As usually happens only in fairy tales, they came up with three strains: one had a shorter than normal periodicity (*per^s*), one had a longer one (*per^l*) and the third was arrhythmic (*perⁿ*). To everybody's amazement, all three strains are alleles — variants of the same gene. Abolishing a behavioural rhythm by mutation would not have been considered a great achievement, but the discovery of a gene involved in the tuning of the cycle time seemed to imply that the gene is close to the mechanism of oscillation, or is even a component of it.

Since this discovery, biochemical, neuro-anatomical and behavioural details about the mutant phenotypes have been elucidated. For instance, a 1-minute rhythm in the wing beat of the courting male is modified in the same manner by the three alleles as are the circadian rhythms. Thus, oscillators with vastly different periodicities seem to be based, at least in part, on

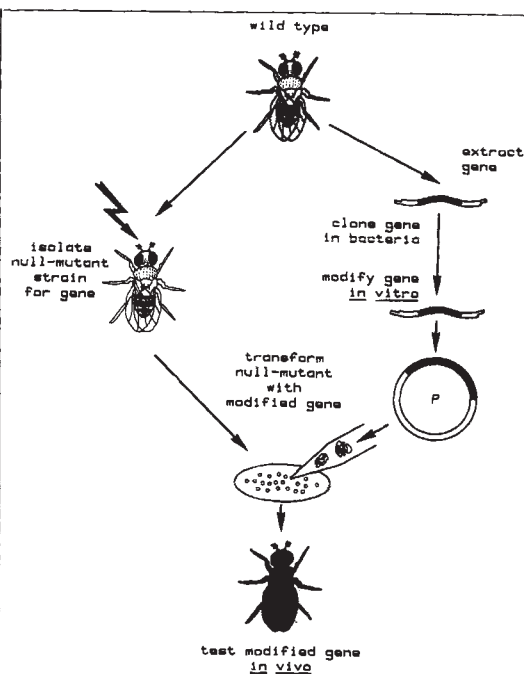
the same mechanism. The gene affects two crucial properties of biological rhythms: their sensitivity to phase-shifting signals and their resistance to changes in temperature. Both findings support the idea that the *per* gene is required in the oscillation processes or in setting up the oscillators during development (see ref. 3).

In the early 1980s, the *per* gene was cloned independently by two groups^{4,5}. The predicted amino-acid sequence of the coding region did not provide any immediate clues as to the function of the protein. Our concepts of clocks are moulded by technical devices such as the pendulum or the sandglass, and we seem to be at a loss in guessing how temperature-compensated biological clocks are built. One hint pointing towards a model of the sandglass type comes from the observation that the amount of *per* gene product is negatively correlated with period length. Extra doses of the wild-type gene shorten the cycle, whereas subnormal expression of the gene elongates it⁶. Thus, *per* might be the sand in the sandglass or a valve regulating its flow.

A benefit of cloning a gene is that immunological and DNA probes can be obtained which allow expression of the gene and its product to be located. In this case, *per* expression is seen in the embryonic nervous system, in the larval salivary gland and in various tissues and cells of the pupa and adult fly^{7,8}. Does this imply that different tissues have their own clocks?

Last year, Young and co-workers at Rockefeller University reported⁹ that gap junctions between the cells of the salivary gland are under the control of the *per* gene. Rhythmicity in this tissue is manifest in the intercellular coupling which is modulated with a circadian periodicity (cited in ref. 3). Young and colleagues propose that the behavioural rhythms are also based on the communication between cells and that the *per* protein exerts its influence at the cell surface or in the extracellular matrix. Whether or not this conjecture is correct, when faced with a complex behavioural phenotype and a complex pattern of gene expression, one would like to attribute particular aspects of one to certain features of the other. Ewer *et al.* in this issue¹ introduce conditional *in vitro* mutants as a new tool for this purpose.

These authors design a temperature-controlled *per* gene by fusing a heat-shock promoter to the *per* gene coding region and transform *per*⁰ flies with this new construct. (Heat-shock proteins are produced in response to stress and are present in most cells.) At high temperatures, these *in vitro* mutants express the *per* gene practi-



In vitro mutagenesis in *Drosophila*. *D. melanogaster* is well suited for the application of this method in behaviour research as the fly combines a rich behavioural repertoire with the following three technical criteria: (1) *Drosophila* genes can be isolated and modified *in vitro*; (2) so-called null mutants (animals without a particular gene) can systematically be found; and (3) a modified gene can be efficiently inserted into the genome of a null mutant.

cally everywhere in the body (as they express heat-shock genes), whereas at low temperature they show no or very little *per* expression. Ewer *et al.* demonstrate that *per* gene expression is necessary only during overt rhythmic behaviour, and that neither the normal developmental profile nor the spatial pattern of expression is essential for the oscillation. One can instantaneously 'switch on' the circadian locomotor rhythm in the adult mutant flies by turning on the gene, and switch the rhythm off as easily by turning the gene off. This result supports (but does not prove) the notion that the *per* gene is part of the oscillation process, but it contradicts those models which require that it is involved in the construction of the oscillator during development.

The function of the *per* gene is still obscure. But *in vitro* mutagenesis should enable *per* and other clock genes to be analysed so that the molecular basis of biological oscillators and their manifestation in behaviour can be understood. □

1. Ewer, J., Roshash, M. & Hall, J.C. *Nature* 333, 82 (1988).
2. Konopka, R.J. & Benzer, S. *Proc. natn. Acad. Sci. U.S.A.* 68, 2112-2116 (1971).
3. Hall, J.C. *Compreh. Insect Physiol. biochem. Pharmac.* 9, 287-373 (1985).
4. Bargiello, T.A. & Young, M.W. *Proc. natn. Acad. Sci. U.S.A.* 81, 2142-2146 (1984).
5. Reddy, P. *et al.* *Cell* 38, 701-710 (1984).
6. Baylies, M.K. *et al.* *Nature* 326, 390-392 (1987).
7. Bargiello, T.A. *et al.* *Nature* 328, 686-691 (1987).
8. Liu, X. *et al.* *Genes Devel.* 2, 228-238 (1988).

Martin Heisenberg is Professor of Genetics at the Institut für Genetik und Mikrobiologie, Universität Würzburg, Röntgenring 11, 87 Würzburg, FRG.

Daedalus

The aerial tadpole

THE hot-air balloon, such an elegant craft in many ways, is a thermodynamic disaster. It burns propane at over 1,000°C, and just lets this temperature decay to 50°C or less, with no attempt to exploit the heat-flux for propulsive power. The snag, of course, is weight. A conventional gas engine would be far too heavy to be lifted by its waste heat discharged into a balloon.

Now Daedalus plans to use the balloon itself as an engine. His idea is to burn fuel in bursts, so that the envelope expands and contracts, ideally in mechanical resonance. This requires a closed envelope, without the conventional opening at the bottom. In the 'expansion' stroke, pressurized propane is released into the envelope through a valved injection burner, entraining air as it enters and burns. In the 'contraction' stroke the valve closes, and the hot air in the balloon cools and leaks out through the carefully controlled porosity of the fabric. Fins on the oscillating envelope could swim it through the sky like a gigantic jellyfish.

But the radial oscillation of a spherical balloon would not propel it very efficiently, even with very big fins. Daedalus's 'aerial tadpole' variant has a much superior mode of oscillation. Its long, cylindrical Zeppelin-type envelope is divided longitudinally into separate chambers, each with its own injection burner.

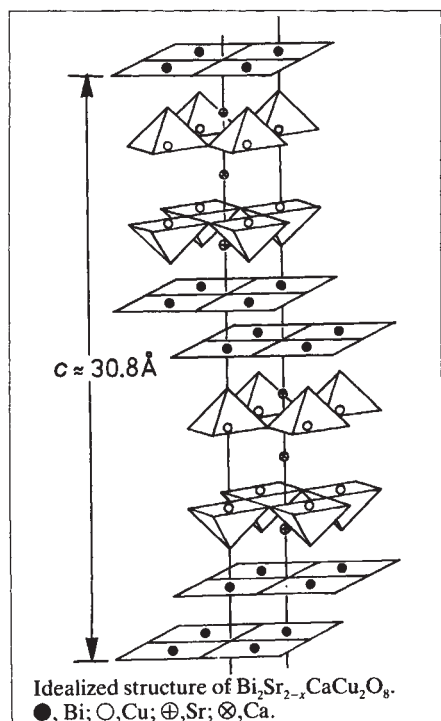
Each chamber is stressed just to buckling by internal sprung and tensioned fibres, so that in the cold the aerial tadpole takes up a zig-zag shape: each chamber is buckled in the middle. But the buckling of an inflated cylinder is very sensitive to small differences of internal pressure. If the burner on any chamber is briefly fired, the sudden rise of pressure straightens that chamber convulsively; inertial overshoot makes it buckle in the opposite direction during the subsequent cooling and contraction. So by firing the burners in proper sequence down the balloon, the sequential straightening and buckling of the chambers will create a lateral travelling wave of large amplitude, a huge resonant tadpole wiggle of the whole balloon, which will propel it with great efficiency.

The aerial tadpole will extend hot-air ballooning far beyond mere recreation. It can travel in any direction, even against the wind: it can hover like a helicopter, while being wonderfully silent, safe, cheap and, of course, delightful to look at. All sorts of short-haul travel, lifting and patrolling and aerial joyriding, will become simple and easy. Aerial tadpoles may even make flying safe enough for individual commuting — making possible the long-predicted traffic-jam in the sky.

David Jones

Control of copper valence in $\text{Bi}_2\text{Sr}_{2-x}\text{CaCu}_2\text{O}_8$

SIR—The recent reports^{1,2} of superconductivity in the system Bi–Sr–Ca–Cu–O have stimulated a great deal of interest, not least because of the relative ease with which these materials might be fabricated into useful forms. According to Chippindale *et al.*³, there are two superconducting phases in the system, with compositions, determined by high-resolution X-ray emission analysis⁴, of $\text{Bi}_2(\text{Sr,Ca})_{2.66}\text{Cu}_2\text{O}_8$



and $\text{Bi}_2(\text{Sr,Ca})_{1.5}\text{CuO}_x$, respectively. Here we are primarily concerned with the first of these phases. The structure of this phase has recently been reported^{5,6}, and indicates a chemical composition that is at variance with the analytical results. According to a combined X-ray and neutron powder study by Bordet *et al.*⁵, the composition is $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ and the space group of the orthorhombic cell ($a=5.401$, $b=5.401$, $c=30.83\text{\AA}$) is *Fmmm* (see figure). The structure reported by Subramanian *et al.*⁶ is broadly similar, although several of the oxygen atoms are not located. It is assigned the composition $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+y}$ and the orthorhombic space group ($a=5.399$, $b=5.414$, $c=30.904\text{\AA}$) *Amma*.

Under normal circumstances, the composition obtained by diffraction methods would be regarded as definitive, but there are several reasons why the chemical results should not be ignored. First, they are corroborated by several other determinations by electron microprobe analysis, including those of Takagi *et al.*⁷, who obtained Bi:Sr:Ca:Cu ratios corresponding to $\text{Bi}_2\text{Sr}_{1.4}\text{CaCu}_2\text{O}_x$, and Takayama-Muromachi *et al.* (preprint), who found

the composition $\text{Bi}_2(\text{Sr,Ca})_{2.50}\text{Cu}_2\text{O}_x$. Zandbergen *et al.*⁸ report similar findings. Second, the two structure determinations of the phase are of modest quality and in entirely different space groups. Neither takes into account the modulation of the structure that leads to an approximately 5-fold incommensurate superlattice along the *b*-axis. The Bordet *et al.* structure⁵ appears to be the most reliable in that all of the atoms, including oxygen, have been located. It does, however, have the curious feature that the composition of the structural model, $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$, would require all of the copper to be present in the +2 oxidation state, even though their bond-strength calculations point to a mean oxidation state of +2.335. The latter is approximately what would be expected by analogy with other superconducting oxides such as $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (ref. 9). In the work of Subramanian *et al.*⁶ the oxygen atoms in the bismuth layers were not located and the composition must therefore be regarded as suspect, but the mean oxidation state of the copper, obtained by bulk chemical analysis, was again found to be approximately +2.33.

How can the conflicting models obtained by analytical and diffraction methods be reconciled? If we assume that the structure reported by Bordet *et al.*⁵ is essentially correct, we can obtain agreement with the analytical data of ref. 3 simply by removing 16 per cent of the strontium atoms from their model. Such a minor error in the structure could easily be accommodated by a slightly higher temperature factor, especially in a structure that is known to have additional subtlety that is not in the model. Removal of this small fraction of strontium yields a composition of $\text{Bi}_2\text{Sr}_{1.68}\text{CaCu}_2\text{O}_8$, in excellent agreement with that reported in ref. 3, which is $\text{Bi}_2\text{Sr}_{1.67}\text{Ca}_{0.99}\text{Cu}_2\text{O}_8$ when the observed Sr/Ca ratio (1.69) is taken into account. This minor change may seem of little importance, but its real significance lies in the fact that it brings the mean oxidation state of the copper up to exactly +2.33, as observed and, indeed, expected. Furthermore, the presence of defects in the strontium layers may be related to the observed modulation of the structure along *b*.

The essence of our proposed model is that the oxidation of Cu^{2+} in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ is controlled by depleting the strontium content according to the general formula $\text{Bi}_2\text{Sr}_{2-x}\text{CaCu}_2\text{O}_8$. This is in contrast to the systems $(\text{La}_{2-x}\text{Ba}_x)\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, in which the copper oxidation state is determined by substitutional defects and oxygen defects, respectively. It is also likely that a similar mechanism operates in the second bis-

moth superconductor, where the analytical data would be consistent with the formulation $\text{Bi}_2(\text{Sr,Ca})_{2-x}\text{CuO}_6$, rather than $\text{Bi}_2(\text{Sr,Ca})_2\text{CuO}_{6+x}$, a possibility that has already been mooted¹⁰.

A.K. CHEETHAM
A.M. CHIPPINDALE
S.J. HIBBLE

*Chemical Crystallography Laboratory,
University of Oxford,
Oxford OX1 3PD, UK*

1. Michel, C. *et al.* *Z. Phys.* **B68**, 421–423 (1987).
2. Maeda, A.H., Tanaka, Y., Fukutomi, N., & Asano, T., *Jap. J. appl. Phys.* **27**, 2 (1988).
3. Chippindale, A.M. *et al.* *Physica C* **152**, 154–156 (1988).
4. Cheetham, A.K. & Skarnulis, A.J. *Analyt. Chem.* **53**, 1060–1064 (1981).
5. Bordet, P. *et al.* *Physica C* (submitted).
6. Subramanian, M.A. *et al.* *Science* **239**, 1015–1017 (1988).
7. Takagi, H. *et al.* *Nature* **332**, 236–238 (1988).
8. Zandbergen, H.W. *et al.* *Nature* **332**, 620–623 (1988).
9. Wu, M.K. *et al.* *Phys. Rev. Lett.* **58**, 908–910 (1987).
10. Torrance, J.B. *et al.* *Solid State Commun.* (in the press).

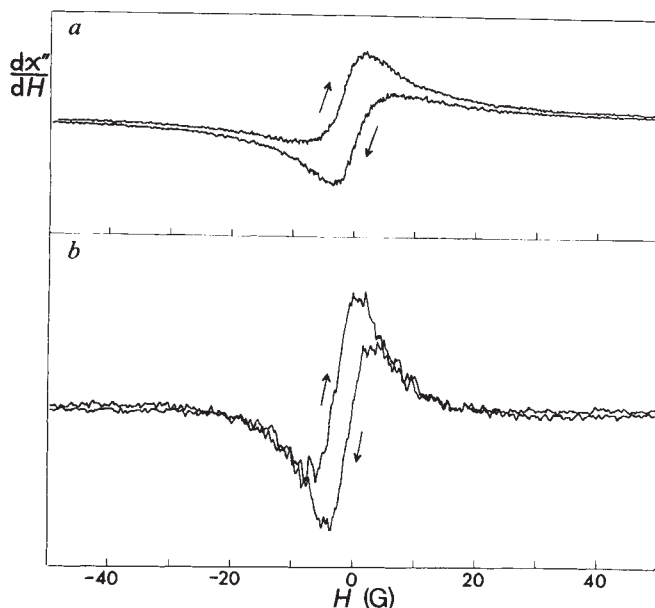
Superconductivity and microwave absorption

SIR—Recently, Sastry *et al.*¹ suggested that the low-field electron spin resonance (ESR) signal of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ superconductors observed with an X-band spectrometer could be attributed to a large zero-field splitting arising from dimerization of electrons on copper pairs in adjacent octahedra. But our work and that of others^{2,3} clearly indicate that the low-field signal is not caused by a resonant effect.

Our work has been done on three ESR spectrometers: S-band (microwave resonance frequency 2.40 GHz); X-band (IBM-Bruker ESR200 at 9.45 GHz); and Q-band (23.0 GHz). For each instrument, field scans through true zero magnetic field were ensured by special circuits. We prepared superconductors from ultrapure Y_2O_3 , BaCO_3 and CuO according to the standard solid-state procedure⁴. The transition temperature is 93 K, with a width less than 2 K.

The figure shows the ESR spectrum obtained from both increasing- and decreasing-field scans at 77 K in the X-band instrument. The breadth of the hysteresis is 5 G, centred at –2 G in the forward direction and 0 G in the backward direction. The breadth of the hysteresis effect depends on the field applied while the sample was cooled through the transition temperature. The spectra obtained with both the lower and higher frequencies are almost identical to those obtained with the X-band instrument. The results at 2.40 GHz are shown in *b* in the figure. Note that the lower frequency is insufficient to induce a transition between

1. Sastry, M.D. *et al.* *Nature* **330**, 49–51 (1987).
2. Blazey, K.W. *et al.* *Phys. Rev.* **B36**, 7241–7243 (1987).
3. Khachatryan, K., Wexler, E.R., Tejedor, P., Stacy, A.M. & Portis, A.M. *Phys. Rev.* **B36**, 8309–8314 (1987).
4. Nelson, D.L., Whittingham, M.S. & George, T.F. (eds) *Chemistry of High-Temperature Superconductors* (American Chemical Society, Washington DC, 1987).



Low-field ESR spectra of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$: *a*, at 77 K, obtained with an X-band spectrometer (microwave frequency 9.45 GHz); *b*, at 60 K, at S-band (2.40 GHz). The measured signal is the field derivative of the imaginary part of the magnetic susceptibility (χ''), as a function of the external magnetic field.

spin states with a zero-field splitting of the order of 0.3 cm^{-1} , as proposed in ref. 1.

Insight into this non-resonant microwave absorption is obtained from one additional measurement, which demonstrates that the amplitude of the absorption is insensitive to whether the H_1 field of the microwave radiation is parallel or perpendicular to the external magnetic field. These observations are consistent with the studies of Blazey *et al.*² and of Khachatryan *et al.*³, and support the proposed mechanism of non-resonant absorption. Blazey *et al.*² used low-field

microwave absorption data to argue that the material is a superconducting glass which consists of weakly coupled superconducting clusters.

T.-S. LIN
L.G. SOBOTKA

Department of Chemistry,
Washington University, St Louis,
Missouri 63130, USA

W. FRONCISZ
National Biomedical ESR Center,
Medical College of Wisconsin,
Milwaukee, Wisconsin 53226, USA

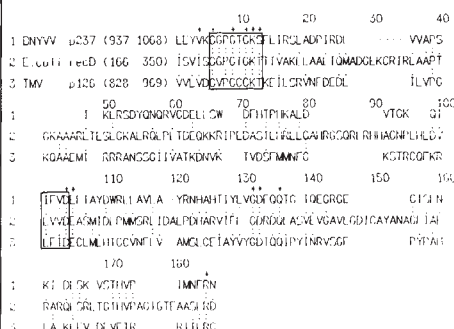
A conserved NTP-motif in putative helicases

SIR—*Escherichia coli* *recBCD* enzyme is a multifunctional protein involved in general recombination which possesses, among other functions, an ATP-dependent DNA helicase activity¹. Inspection of the amino-acid sequences of the complex subunits reveals that two of them, *recB* and *recD*, contain a consensus pattern of residues characteristic of the catalytic sites of many enzymes that use nucleoside triphosphates (NTPs), the so-called NTP-motif². This prompted us to perform a more detailed computer-assisted comparison of the sequences of these two proteins with those of other enzymes of this class.

This comparison reveals an unexpected similarity between *recD* and the NTP-motif-containing domain of a non-structural protein of beet necrotic yellow vein virus (BNYVV), a positive-strand RNA plant virus. This domain belongs to a recently identified family of homologous viral proteins (domains) involved in virus RNA replications^{3,4}. In these proteins the NTP-motif is one of the most strictly conserved stretches of sequence.

For one of them, tobacco mosaic virus (TMV) protein p126, the NTP-binding capacity has been demonstrated experi-

mentally⁵. The optimal alignment of the NTP-motif-containing domains of *recD* and of the BNYVV and TMV proteins is shown in the figure. The similarity is high enough to suggest a monophyletic origin for the compared domains (see table).



Optimal alignment of the NTP-motif-containing domains of *recD* and the presumptive NTPases of BNYVV and TMV. Sequences from refs 1 (*recD*), 7 (BNYVV) and 8 (TMV). Alignment generated by the program OPTAL, based on the original algorithm of Sankoff⁶. Identical residues (:) and conservative replacements (.) are highlighted. Asterisks, residues constituting the putative viral NTP-motif consensus^{4,11}. The residues constituting the NTP-motif proper are boxed.

Summary of the alignments

	<i>recD</i>	BNYVV	TMV
<i>recD</i>	—	28.8(51.5)	16.2(35.2)
BNYVV	5.9	—	20.5(31.8)
TMV	3.7	3.2	—

Below the diagonal: Alignment scores calculated in standard deviation (s.d.) units for the three alignments. Note that the similarity between the *recD* and BNYVV segments is the highest and exceeds the threshold of 5 s.d. thought to be indicative of a true evolutionary relationship¹⁰. Above the diagonal: per cent similarity expressed as strict coincidence and, in parentheses, coincidence including conservative replacements.

Strikingly, the similarity between the *recD* and BNYVV sequences is even higher than that between the two viral sequences. Such a pronounced sequence similarity could be due to conservation of a specific NTP-requiring function. Thus, viral NTP-motif-containing proteins may be subunits of RNA helicases involved in the unwinding of double-stranded replication forms during viral RNA replication, and in recombination between RNA genomes, a process recently described for plant viruses⁸.

ALEXANDER E. GORBALENYA
EUGENE V. KOONIN
ALEXEI P. DONCHENKO
VLADIMIR M. BLINOV

Institute of Poliomyelitis and Viral
Encephalitis,
USSR Academy of Medical Sciences,
142782 Moscow Region, USSR

1. Finch, P.W. *et al.* *Nucleic Acids Res.* **14**, 8583–8594 (1986).
2. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. *EMBO J.* **1**, 945–951 (1982).
3. Gorbalenya, A.E., Blinov, V.M., Koonin, E.V. *Molekul. Genetika* **11**, 30–36 (1985).
4. Gorbalenya, A.E. *et al.* *Molekul. Biol.* **21**, 1566–1571 (1987).
5. Evans, R.K., Haley, B.E. & Roth, D.A. *J. biol. Chem.* **260**, 7800–7804 (1985).
6. Bujarski, J.J. & Kaesberg, P. **321**, 528–531 (1986).
7. Bouzoubaa, S. *et al.* *J. gen. Virol.* **68**, 615–626 (1987).
8. Golet, P. *et al.* *Proc. natn. Acad. Sci. USA* **79**, 5818 (1982).
9. Pozdnyakov, V.I. & Pankov, Yu. A. *Int. J. Peptide Protein Res.* **17**, 284–291 (1981).
10. Doolittle, R.F. *Science* **214**, 149–159 (1981).
11. Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. & Koonin, E.V. *J. molec. Evol.* (in the press).

A new superfamily of replicative proteins

SIR—I report a set of 21 related proteins, identified by computer searches, which are all involved in nucleic-acid replication and/or recombination. These include two *Escherichia coli* ATP-dependent helicases, four essential human herpesvirus proteins (probably also helicases), two exonuclease V (ExoV) subunits, and the yeast PIF protein (involved in mitochondrial DNA recombination). These helicases and nucleases are structurally related to a set of conserved domains (20–70% amino-acid identity) that are common in RNA viruses¹, and which I used as the starting point for the search.

The *uvrD* and *rep* helicases are both

Motif	I	II	III	IV	V	VI	Source						
uvrD	26	VLGAGSGKTRVLV	174	NILVDEFQNTN	16	VMIVGDDQSIY	26	QNYRSTSI	267	IMTLHS-AGKL-EFPQVIVG	23	LAYGVVTRA	(2)
rep	19	VLGAGSGKTRVIT	175	YLLVDEYQDTN	16	FTVVGDDQSIY	26	QNYRSGRI	271	IMTLHA-SKGL-EFPYVVMV	22	LAYGVITRA	(2)
recB	20	IFASAGTKTETIA	345	VAMIDEQDTH	18	LLIGSPQKAIY	24	TWRSAPFM	286	IVTIHK-SKGL-EYPLVWLF	44	LLYVALTRS	(4)
recD	164	ISGGPGTKTTTVA	82	VLVDEASMD	16	VIFLGRDQLAS	24	QLSRLTGH	198	AMTVHK-SQGS-EFDHAALIL	11	LVYTAVTRA	(21)
EBV	69	ITGTAGAGKSTSVS	113	VIVWDEAGTLS	26	IVCVGSPTQDA	44	NNKRCITDVQ	426	AMTIK-AQGL-SLNKVAICF	9	HVYVALSRA	(22)
HCMV	117	VTGTAGAGKSTSIQ	125	IIVIDEGLML	26	IICVGSPTQTEA	44	NNKRCITDLD	513	AMTIK-SQGL-SLEKVADEF	10	HVYVALSRV	(6)
HSV	94	ITGNAGSGKSTCVQ	136	VIVIDEAGLLG	26	LVCVGSPTQTAS	44	NNKRCVEHE	442	AMTIR-SQGL-SLDKVAICF	8	SAYVAMSRT	(6)
VZV	87	ISGNAGSGKSTCIQ	135	VIVIDEAGLLG	26	IICVGSPTQTDS	44	NNKRCQEDD	447	AMTIAR-SQGL-SLEKVAICF	8	SVYVAMSRT	(23)
PIF	255	YTGAGTGKSLILR	46	ALVWDEISMLD	25	LIFCGDFQLFP	29	KVFRQGDV	219	MQTHQNSAGKRLRILRVFKA	33	QAYVALSRA	(10)
ALMV	821	VDVGAGCGKTTNIK	55	RLIFDECFLOH	15	VIGFGDTEQIPF	22	ITWRSPADA	66	IFTTHE-AQCK-TFDNVYFCR	19	NGLVALSRH	(1)
BMV	687	VDVGAGCGKTTAIK	54	RLVWDEAGLLH	15	VLAGDTEQISF	22	KTYRCQDV	78	IKTVHE-AQGI-SVDNVTILVR	13	YCLVALTRH	(1)
CMV	709	VDVGAGCGKTTAIK	54	RLVWDEAGLLH	15	ALCPGDSQIAF	22	TTFRSPQDV	79	IKTVHE-SQGI-SEDHVTILVR	13	YCLVAVTRH	(1)
TMV	829	VDVGAGCGKTTKEIL	57	RLIFDEGLMLH	15	AYVGDTQQIPY	24	TLTRCPADV	62	VHTVHE-VQGE-TYSDVSLVR	14	HVLVALSRH	(1)
ToMV	829	VDVGPGCGKTTKEIL	57	RLIFDEGLMLH	15	AYVGDTQQIPY	24	TLTRCPADV	62	VHTVHE-VQGE-TYADVSLVR	14	HVLVLSLRH	(24)
TRV	901	VDVGPGCGKSTMIY	56	VHLEDEALMAH	15	CICQGDQNIQSF	24	EYTRSPADV	64	VSTVHE-SQGE-TFKDVLVLR	13	YLVVALSRH	(1)
SFV	183	VGVFGSGKSAIHK	50	ILYVDEAFACH	16	VVLGDPKQCGF	18	ISRCRTPQV	58	VMTAAA-SQGL-TRGVYAVR	14	HVNVLTRT	(1)
SV	183	VIGTPGSGKSAIHK	50	VLVYDEAFACH	16	VVLGDPKQCGF	24	ISRCRTPQV	58	VMTAAA-SQGL-TRKGVYAVR	14	HVNVLTRT	(1)
IBV	1209	VQGGPGSGKSHFAI	54	ILLVDEVSMILT	15	VVYVGDPAQLPA	30	KCYRCPEKI	82	VQTVDS-SQGS-EYDYVIFCV	11	RFNVALTRA	(25)
BNYV1	893	VKGPGTGKSFILR	48	IIFVDEFTAYD	11	IYLVGDEQQGTI	25	MNFRNPHD	72	KITVRA-NQGS-TYDNVVLV	12	LNLVALSRH	(26)
BNYV2	121	VLGAGVGKSTSIK	49	ITLVDEVTIRVH	11	VICFGDPAGQLN	18	ASRREGKAT	67	SILYD-AHQG-TYDVTIIL	13	VRVALLTRA	(26)
BSMV2	267	ISGVFGSGKSTIVR	41	LLIIDEYTLAE	11	VLLVGDVAQGLA	18	TTYRLQET	62	CALAIQ-VQGE-EFDSVTILF	12	LRLVALSRH	(27)
		* * * *	*****	*****	*****	*****	*****	*****					
	V G A G K S	V D E	G D Q	R	T S Q G E V	V	V A L S R						
	I A P T	I	S		L A K S A	A	T L V T						
	Y	F			V H T		G M						
					N A R		S I						

Alignment of conserved motifs from six protein families (single-letter code). Hyphens, gaps introduced during alignment. The number of residues between each motif is shown. At the base of alignment, characteristic residues have been shown where there are less than five alternatives. The main RNA virus family includes: ALMV, alfalfa mosaic virus 1a; BMV, brome mosaic virus 1a; CMV, cucumber mosaic virus 1a; TMV, tobacco mosaic virus (common strain) p126; ToMV, tobacco mosaic virus (tomato strain) p126; TRV, tobacco rattle virus p134; SFV, Semliki Forest virus nsP2; SV, Sindbis virus nsP2. The last two are alphaviruses infecting animals. Conserved regions in this protein family were used to construct motifs for the program TPT¹⁴, which extracted uvrD from the Protein Information Resource database. A general database search using FASTP¹⁵ then matched a region from recD with Epstein-Barr virus (EBV) BBLF4. Related genes in other herpesvirus: human cytomegalovirus (HCMV) PS3 (J.A. Martignetti, personal communication), herpes simplex virus (HSV) UL5, and varicella-zoster virus (VZV) gene 55, are strongly conserved. Similarities between the yeast PIF sequence⁶ and uvrD, and between infectious bronchitis virus (IBV) F2, beet necrotic yellow vein virus (BNYV1) 237K protein and the alphavirus nsP2 family were noted by the original authors. The relationship between the (BNYV2) 42K and barley stripe mosaic virus (BSMV2) 58K proteins was also noted, but their similarity to all the other proteins was not. Asterisks, positions used for statistical analysis¹⁶. The probability of each motif occurring by chance was $\sim 10^{-6}$, and for the six- and seven-member patterns 1.9×10^{-33} and 1.6×10^{-39} , respectively.

DNA-dependent ATPases sharing 37% amino-acid identity². The recB and recD are subunits of ExoV. The former probably carries out the helicase activity of the ExoV holoenzyme because it has been shown to bind ATP³ and single-stranded DNA⁴; also, it has 20% amino-acid identity with the above helicases. Its extra sequences presumably carry out ExoV-specific functions. The recD does not have a demonstrated helicase function, binds ATP only in the presence of the other subunits³, and has diverged further than the others. The structural resemblances between these proteins strongly suggest that they belong to a DNA helicase family that is distinct in evolutionary terms from other *E. coli* helicases, although they are related to PIF, a yeast nuclear gene product involved in mitochondrial DNA repair and recombination⁵.

Genetic data in herpes simplex virus⁶ show that UL5 is essential for DNA replication, a finding that is consistent with its stringent conservation among other herpesviruses. UL5 could also be a helicase⁶, a function possibly common to all groups within the superfamily, although the level of similarity between the groups is too low to allow a direct functional comparison. The available evi-

dence shows that the RNA viral protein domains containing these conserved motifs are involved in replication⁷, but are distinct from the suspected polymerase domain⁸.

The very broad occurrence of these motifs in otherwise unrelated proteins strongly suggests that they all carry out crucial nucleoside triphosphate-dependent steps in nucleic acid replication. Motif I is well-known⁹, and is found almost exclusively in ATP and GTP binding proteins. Where the crystal structure is known¹⁰⁻¹², the motif forms a loop which binds one of the phosphates, another being bound by an aspartate via a magnesium ion. There is a suitably located, strictly conserved aspartate in motif II. Motif III was previously noted¹³ to resemble a conserved region from viral DNA polymerases. This study shows that the proteins have no overall relationship, though it remains possible that they have similar local folds. Note also that only the presumed DNA binding proteins have tyrosine at position 3 of motif VI and another motif between motifs I and II.

These motifs probably represent common secondary structures that make up functional sites for pyrophosphate, magnesium or nucleic acid binding, while the rest of the sequence provides the struc-

tural framework, the specificity and any supplementary functions. The resolution of the exact structure, function, and evolutionary significance of these motifs is still awaiting further crystallographic analysis.

T.C. HODGMAN

MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge, CB2 2QH, UK

- Hodgman, T.C. & Zimmern, D. in *RNA Genetics* (eds Holland, J. et al.) (CRC Press, Boca Raton, in the press).
- Gilchrist, C.A. & Denhardt, D.T. *Nucleic Acids Res.* **15**, 465-475 (1987).
- Julin, D.A. & Lehman, I.R. *J. biol. Chem.* **262**, 9044-9051 (1987).
- Finch, P.W. et al. *Nucleic Acids Res.* **14**, 8573-8582 (1986).
- Foury, F. & Lahaye, A. *EMBO J.* **6**, 1441-1449 (1987).
- McGeoch, D.J. et al. *J. Virol.* **62**, 444-453 (1988).
- Fuller, F.J. & Marcus, P.I. *Virology* **107**, 441-451 (1980).
- Kamer, G. & Argos, P. *Nucleic Acids Res.* **12**, 7269 (1984).
- Walker, J.E. et al. *EMBO J.* **1**, 945-951 (1982).
- Fry, D.C. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 907 (1986).
- La Cour, T.F.M. et al. *EMBO J.* **4**, 2385-2388 (1985).
- Jurnak, F. *Science* **230**, 32-36 (1985).
- Hodgman, T.C. *Nucleic Acids Res.* **14**, 6769 (1986).
- Bashford, D. et al. *J. molec. Biol.* **196**, 199-216 (1987).
- Lipman, D.J. & Pearson, W.R. *Science* **227**, 1435 (1985).
- Staden, R. *CABIOS* **4**, 53-60 (1988).
- Ohno, T. et al. *J. Biochem.* **96**, 1915-1923 (1984).
- Boursnell, M.E.G. et al. *J. gen. Virol.* **68**, 57-77 (1987).
- Bouzoubaa, S. et al. *J. gen. Virol.* **68**, 615-626 (1987).
- Gustafson, G. & Armour, S.L. *Nucleic Acids Res.* **14**, 3895-3909 (1986).
- Finch, P.W. & Emmerson, P.T. *Nucleic Acids Res.* **12**, 5789-5799 (1984).
- Finch, P.W. et al. *Nucleic Acids Res.* **14**, 8583 (1986).
- Baer, R. et al. *Nature* **310**, 207-211 (1984).
- Davison, A.J. & Scott, J.E. *J. gen. Virol.* **67**, 1759 (1986).

Lifting heavy loads not just by the Egyptians

SIR—The astute suggestion made by John Cunningham (*Nature* **332**, 22-23; 1988) that certain old Egyptian art work illustrates a method of raising heavy weights prompted me to look into that compendium of primitive clever invention, Francis Galton's *Art of Travel* (5th edn, David and Charles, Newton Abbot, 1971).

Sure enough, two related systems, which make use of this principle of successive small forces being applied to a series of springy supports, are mentioned under the heading 'Accumulation of Efforts'.

To lift and swing forward a heavy log in a tropical forest, 'the labourer gets hold of one of these creepers that runs from the top of the boughs of a tree in the direction in which he wants to move his log, and pulling this creeper home with all his force, bending down the bough, he attaches it to the log; then he goes to another creeper and does the same with that; and so on until he has accumulated strain of many bent boughs, urging the log forward and of sufficient power to move it'. The other example concerns a commercially available 'accumulator', consisting of cords of india-rubber each hooked to a fixed ring at one end, the other ends being then hooked one by one to the object to be moved.

W. RIGBY

2 Sweden Park,
Ambleside, Cumbria LA22 9EY, UK

Primitive ancestry of transfer RNA

SIR—Weiner and Maizels propose that primitive genomes were RNA strands capped by transfer (t)RNA-like sequences serving as telomeres and replication initiation sites¹. Central to this 'genomic tag' model, which has been discussed in *News* and *Views*², is the observation that some viral genomes have 3'-terminal tRNA-like structures that Weiner and Maizels suggest are direct descendants of the earliest tRNAs, implying that the primitive tRNA was a continuous molecule. A subset of tRNAs in biological systems as diverse as archaeobacteria, chloroplasts, fungi and mammalian cells, however, is known to be interrupted by introns. Their broad occurrence and conserved location in anticodon loops suggest that introns existed in tRNA when the primary lines of descent were being established, thus raising the question of the role of the intron in a prebiotic milieu. Present knowledge of tRNA introns offers few clues.

Unlike most tRNA introns, those of chloroplast tRNAs are up to 2.5 kilobases long and can adopt structures similar to autosplicing introns^{3,4}. Although these introns differ from those of nuclear tRNAs they raise the possibility that nuclear tRNA introns descended from structures similar to those in plastids. This seems plausible in view of the strikingly diverse tRNA splicing pathways of yeast⁵, *Hela* cells⁶ and halobacteria⁷. The diversity can be accounted for by supposing that primitive tRNA introns were originally capable of self-splicing, and that independent pathways evolved as intron excision became protein-mediated, whereas nuclear tRNA introns were concomitantly pared down to the minimal sizes observed today.

Weiner and Maizels also propose that a variant of a self-splicing ribozyme had the capacity to aminoacylate tRNA¹. I suggest that the intron of the primitive tRNA, when spliced out, served as this rudimentary ribozymic aminoacyl-tRNA synthetase.

Several interesting consequences follow from the 'internal ribozymic synthetase' model. First, co-synthesizing tRNA with its cognate ribozymic synthetase ensures that the enzyme level matches that of substrate. Interestingly, although present-day protein aminoacyl-tRNA synthetases act catalytically, the *in vivo* molar ratios of each synthetase to cognate tRNA are surprisingly high (0.05–1.0 in *Escherichia coli*⁸) and near the equimolar amounts implied by the model. Co-synthesis of the two molecules would promote stability of a cognate tRNA-synthetase complex by overcoming diffusional and entropic barriers, and would thereby minimize mischarging between non-cognate pairs.

If primitive replication was not efficiently processive, an additional advantage of locating the synthetase within the cognate tRNA would be that both molecules are produced from a single precursor transcript. Thus, abortive replication would produce an incomplete transcript incapable of splicing out the synthetase, so that no functional molecules are produced and the equimolar balance between synthetase and tRNA is maintained. By contrast, were the tRNA and synthetase units unlinked, an inefficient replicase would abort the longer synthetase transcript more frequently, producing fewer synthetases per cognate tRNA. The result would be less efficient charging and also possibly the promotion of misacylation.

Finally, the internal synthetase model offers insights into the intriguing involvement of protein aminoacyl-tRNA synthetases in fungal mitochondrial RNA splicing^{9,10}. Primitive organisms in which tRNA splicing and aminoacylation were coupled may have evolved proteins mediating both reactions and may therefore have recruited protein synthetases to splice structurally related introns in other interrupted genes.

Because tRNA maturation and aminoacylation are separated by the nuclear envelope in eukaryotes, it is doubtful whether vestiges of direct interactions

between splicing and aminoacylation remain in nucleated organisms. More promising is the study of RNA processing in intron-containing systems lacking the nuclear-cytoplasmic division, such as organelles and archaeobacteria. In view of the structural similarities of chloroplast tRNA introns to autosplicing and fungal mitochondrial introns, it is conceivable that tRNA processing in such systems involves other examples of intron self-excision and components shared between splicing and aminoacylation. In that case the primitive tRNA may not only have been self-splicing, but 'self-charging' as well.

CALVIN K. HO

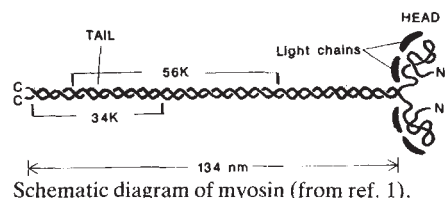
*Division of Biology,
Californian Institute of Technology,
Pasadena, California 91125, USA*

1. Weiner, A.M. & Maizels, N. *Proc. natl. Acad. Sci. U.S.A.* **84**, 7383–7387 (1987).
2. North, G. *Nature* **328**, 18–19 (1987).
3. Bonnard, G., Michel, F., Weil, J.H. & Steinmetz, A. *Molec. gen. Genet.* **194**, 330–336 (1984).
4. Shinozaki, K., Deno, H., Sugita, M., Kuramitsu, S. & Sugiura, M. *Molec. gen. Genet.* **202**, 1–5 (1986).
5. Greer, C.L., Peebles, C.L., Gegenheimer, P. & Abelson, J. *Cell* **32**, 537–546 (1983).
6. Filipowicz, W., Konarska, M., Gross, H.J. & Shatkin, A.J. *Nucleic Acids Res.* **11**, 1405–1418 (1983).
7. Daniels, C.J., Douglas, S.E. & Doolittle, W.F. *System. Appl. Microbiol.* **7**, 26–29 (1986).
8. Jakubowski, H. & Goldman, E. *J. Bact.* **158**, 769 (1984).
9. Akins, R.A. & Lambowitz, A.M. *Cell* **50**, 331–345 (1987).
10. Herbert, C.J., Labouesse, M., Dujardin, G. & Slonimski, P.P. *EMBO J.* **7**, 473–483 (1988).

When the left hand doesn't know. . .

SIR—The structure of myosin depicted in a recent *News* and *Views* article¹ is not correct. The figure properly represents the myosin tail as a fibrous coiled-coil of the α -helical carboxy termini of the myosin heavy chains. But the sense of the coiled-coil is not right-handed, as indicated in that figure, but is in fact left-handed.

The coiled-coil structure is left-handed precisely because the 3.6-residue per turn



α -helix is right-handed. In 1952, Crick² proposed the 'knobs-into-holes' packing scheme for polypeptide helices and postulated that α -keratin is a coiled-coil. His analysis indicated that the small interhelix angle necessary for a coiled-coil requires that it has a rotational sense opposite to that of the α -helices. It is now evident that this structure describes myosin and many other fibrous proteins (see ref. 3).

The oversight is forgivable (I have made the same error⁴) because, even after 36 years, the basic structural features of the

coiled-coil are misrepresented in the educational biochemical literature. I have examined the treatment of fibrous proteins in biochemistry textbooks in the United States. About half do not show molecular detail, whereas some show a coiled-coil with no specific helix sense. Of five popular texts that do indicate a helical sense for the two-strand coiled-coil of either myosin or tropomyosin, all show it as right-handed. One text also incorrectly depicts α -keratin as a right-handed coiled-coil. Two recent articles in *Scientific American* also erroneously depict α -helical coiled-coils as right-handed. However, as none of the authors specifically describes right-handed α -helical coiled-coils, these errors may merely reflect oversight and the predispositions of (mostly) right-handed graphic artists.

The coiled-coil α -helical motif appears to describe the structure of a diverse group of proteins that includes α -keratin, fibrinogen, tropomyosin, paramyosin and some silk proteins, as well as myosin. Perhaps it is time that we all get it right (sorry, left).

RICHARD D. LUDESCHER

*Department of Chemistry,
Wichita State University,
Wichita, Kansas 67208, USA*

1. Sheetz, M.P. *Nature* **331**, 212–213 (1988).
2. Crick, F.H.C. *Nature* **170**, 882–883 (1952).
3. Fraser, R.D.B. & MacRae, T.P. *Conformation in Fibrous Proteins* (Academic, New York, 1973).
4. Ludescher, R.D., Eads, T.M. & Thomas, D.D. in *Optical Studies of Muscle Cross-Bridges* (eds Baskin, R.J. & Yeh, Y.) 33–66 (CRC, Boca Raton, Florida, 1987).

The importance of being Ernst

Philip Kitcher

Toward a New Philosophy of Biology: Observations of an Evolutionist. By Ernst Mayr. Harvard University Press: 1988. Pp.564. \$35. To be published in Britain in June.

THE genre is as familiar as it is disheartening. A scientist, having earned the acclaim of the learned world for distinguished scientific work, more or less lightly turns to thoughts of philosophy, offering the public a collection of observations. Embarrassed reviewers struggle to mix their respect for the years of pioneering research with vague and muted worries about the banality of the latest product. With luck, the ventures in philosophy vanish without a trace, neither receiving the serious attention that they do not deserve nor dimming the lustre of the genuine scientific achievements.

Ernst Mayr is an exception. For decades he has campaigned, with his usual vigour, against those whose horizons in the philosophy of science are bounded by the grand successes of twentieth-century physics. In the process, he has assisted, as biological midwife-in-chief, at the birth of the 'new philosophy of biology' celebrated in the title of his latest collection. Biological problems and biological theories have become the focus of philosophical attention, no longer to be trimmed and squeezed into the philosophical moulds that have been crafted from studies of 'real' science (i.e. physics).

Toward a New Philosophy of Biology is, in part, a manifesto for a revolution that has already happened. Mayr-watchers will find few surprises in the themes that are sounded. Mayr has included essays arguing for the autonomy of biology from physics and chemistry, inveighing against reductionism in general and 'bean-bag genetics' in particular, urging his favoured separation of teleological concepts, pleading the case for the concept of the biological species, and emphasizing the importance of 'population thinking'. Nor will the *cognoscenti* be astonished by the fact that these themes are repeated, reiterated, rephrased, redefended, recapitulated — and, just in case you missed the point, hammered home again.

Mayr's collection has defects that are so obvious that they may prevent a profitable reading. Although there is considerable redundancy, little has been done to integrate the essays. The treatment of other people's ideas is often idiosyncratic. Mayr

will fasten on a phrase from a book or essay — something that strikes his fancy or provokes his reprimand — wrenching it from the context of a different set of ideas and, not infrequently, radically misinterpreting it. There is little systematic discussion of opposing positions or dissection of



Life in style — a design from Mayr's book.

arguments against Mayr's favoured views. Moreover, in many cases, the primary concern seems to be to set the record straight, to explain that Mayr did not mean what his critics have taken him to mean or to argue that certain contemporary doctrines stem from his own much earlier formulations.

Viewed from the right perspective, these shortcomings fade into the background. What Mayr has to offer here, as in his earlier books, is nothing less than a vision of biology that places neodarwinian evolutionary theory firmly at the centre. Insisting on the distinction between questions about proximate causal factors (inquiries into mechanisms) and questions about 'ultimate' causal factors (evolutionary studies), he starts from the pheno-

menon of organismic diversity, seeing the understanding of the enormous variety of living things as the fundamental task of the biological sciences. The first goal is to find an appropriate way to map the diversity, and so, for Mayr, systematics, far from being a peripheral discipline for lapsed philatelists, is central to the biological enterprise. Debates among protagonists of rival classificatory systems, lucidly surveyed in the essay "Toward a Synthesis in Biological Classification", are not irrelevant quarrels but disputes affecting the ways in which all biological questions are posed. In attempting to resolve these debates, Mayr pursues a characteristic strategy, attempting to fashion an eclectic position that will accommodate what he perceives as the insights of rival points of view. That view is defended with an extraordinary breadth of knowledge of the diversity of life. One may quarrel with the interpretations and the arguments, but Mayr's power to discern biological connections and also to identify the telling example should excite unqualified admiration.

Evolutionary explanation of the diversity of life goes hand in hand with the proper description of that diversity. The main outlines of Mayr's preferred explanation are familiar. Species are divided from one another by the criterion of reproductive isolation, and a principal project of evolutionary explanation is to specify the processes through which reproductive isolation is attained. On Mayr's account, an important part of the story is that geographical separation should precede reproductive isolation. One scenario that he takes to be played out with considerable frequency is that speciation is initiated by the splitting off of a peripheral population of a species, in which there *may* be an unusual

distribution of alleles and which *may* face atypical ecological challenges. Evolving under natural selection (as the main, but not the only, evolutionary 'force'), this population founds a lineage in which descendant populations are reproductively isolated from populations of the parental species.

Toward a New Philosophy of Biology is principally concerned with articulating this explanatory project and with the concepts and questions involved in it. Mayr gives less emphasis to exhibiting the positive power of his approach — although there are illuminating essays on classification in birds and on speciation in fish — and spends most of his pages in combat with a many-sided opposition. The issues covered include all the principal contro-

versies about neodarwinism that have figured in the recent literature. To what extent is natural selection the major force in evolutionary change? Has neodarwinism committed itself to an unwarranted adaptationist programme? Is the unit of natural selection the gene? Are appeals to higher-level processes (group selection, species selection) coherent? Are they needed to make sense of the diversity of life? Can speciation occur without prior geographical isolation? Is there good evi-

Harvard University

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Ernst Mayr — vigorous campaigner.

dence for the theory of punctuated equilibrium? Does that theory pose a serious challenge to neodarwinism? Does neodarwinism need to be expanded, revised, reformulated — or simply scrapped?

Mayr's responses to these questions are typically complex and learned, building a many-sided version of neodarwinism that belies the caricatures of its most vocal critics. Sometimes one has the impression that the position is too flexible, that in the hallowed manner of Church of England theology and of Walt Whitman at his most embracing, the pointed criticisms are being enveloped in an enormous cocoon. Nonetheless, even though Mayr makes no sustained attempts to analyse the serious arguments of the multifarious dissenters, the construction of his own version of neodarwinian orthodoxy does valuable service by showing in detail just what the target is. If Ernst Mayr's articulation of the synthesis did not exist, then, like God, it would have to be invented.

In terms of analytical depth, Mayr's essays have already been superseded by work that he has partly inspired. In breadth of vision, his philosophical observations on biology in general and neodarwinism in particular are rivalled only by his own earlier work. *Toward a New Philosophy of Biology* is a book to be developed, to be argued with, a book whose margins should be filled with excited scribbles. Philosophy of biology does not stop here, but, as Mayr's title indicates, this is a place to start. □

Philip Kitcher is a Professor in the Department of Philosophy, University of California at San Diego, La Jolla, California 92093, USA.

Feeling under the weather

Michael Shepherd

Hypochondria: Woeful Imaginings. By Susan Baur. *University of California Press: 1988. Pp.252. \$19.95, £10.*

ACCORDING to the dust-jacket of this book, the author is completing a doctorate in counselling psychology and has written a prize-winning account of the history of oceanography. Interesting as these achievements may be, neither of them would seem to qualify her to provide "a comprehensive understanding of hypochondria", a topic on which she appears to have published no previous work.

Ms Baur's bibliography covers the anglophonic medico-psychiatric literature reasonably well, along with works relating to a number of famous hypochondriacs — Samuel Johnson, Charlotte Brontë, Leo Tolstoy, Sara Teasdale and Alice James among them — and prominent fictional characters, including Molière's Argan and Smollett's Matthew Bramble. From this material she has fashioned a pastiche, juxtaposing summaries of clinical and sociological studies with real-life anecdotes, literary references, biographical details and her own reflections. Individual chapters are devoted to such specific aspects of the hypochondriacal syndrome

as symptoms, age, occupation, treatment, doctors and the family, and the author skips nimbly from one issue to another, barely indicating whether she is discussing a disease state, a social response or an aspect of the human condition.

For whom is a book of this type intended? It will surely seem too superficial for physicians and too pretentious for many non-medical readers. Ms Baur's appeal may be principally to sympathizers with the chapter entitled, "Hypochondria and Our Cultural Values". The cultural values turn out to be largely those of certain segments of modern American society where she finds "a pervasive preoccupation with health and body image" and "a tendency to retreat from the complexities of an unmanageable world and to project our pessimism as well as our managerial propensities onto our bodies instead". This is a world of personal angst, a world that is depicted with more precision, style and humour by Woody Allen:

I believe my consumption has grown worse. Also my asthma. The wheezing comes and goes, and I get dizzy more and more frequently. I have taken to violent choking and fainting. My room is damp and I have perpetual chills and palpitations of the heart. I notice, too, that I am out of napkins. Will it ever stop?

Susan Baur doesn't seem to think it will. □

Michael Shepherd is Professor of Epidemiological Psychiatry at the Institute of Psychiatry, University of London, Denmark Hill, London SE5 8AF, UK.

Per ardua ad astra

Jack Meadows

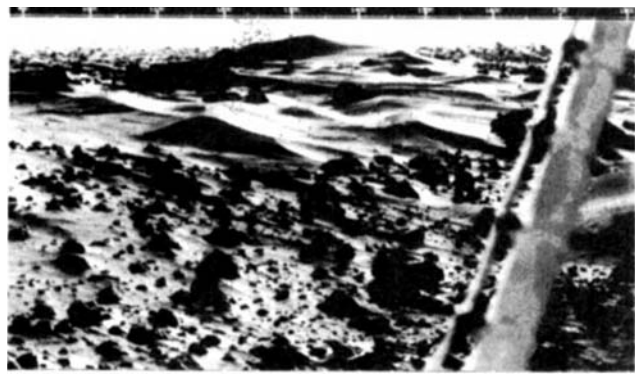
History of the Royal Astronomical Society. Vol.1, 1820–1920, edited by J. L. E. Dreyer and H. H. Turner. Vol.2, 1920–1980, edited by R. J. Tayler. *Blackwell Scientific: 1987. Vol.1 pp.258, hbk £35, pbk £14.50. Vol.2 pp.262, hbk £29.50, pbk £14.50.*

FROM the late eighteenth century onwards, specialist scientific societies sprang up in London. They were typically formed by like-minded gentlemen who came together to discuss and develop their hobby. Thus John Herschel's diary for 12 January, 1820, records: "Dine at the Freemason's Tavern to meet Dr. Pearson and other gentlemen to consider forming an Astronomical Society".

In retrospect, such specialization seems inevitable. For over a century previously, the Royal Society had covered all branches of science, but the growth of scientific knowledge had now reached the stage where individuals found it easier to concentrate on a limited area. So specialist societies appeared. At the time,

the need for change seemed less obvious. The then president of the Royal Society, Sir Joseph Banks, saw no reason for new societies. He persuaded the first proposed president of the Astronomical Society, the Duke of Somerset, not to accept. As one of the original secretaries noted: "a similar, and indeed a more violent, attack was made at the establishment of the Geological Society, and also of the Royal Institution, and which only tended to unite more firmly the original members".

The happy outcome of this quarrel was that Sir William Herschel became the society's first president. The Royal Astronomical Society (RAS), as it became in 1831, has since adopted a drawing of his telescope as its insignia. William Herschel, by then an old man, was president in name rather than practice. His son, John, was involved in the running of the society for many years, becoming president on three occasions. Such frequent office-holding was common in the early decades of the society's history. It partly reflected the society's origins in a small group, and partly arose because only a small number of Fellows then devoted themselves full-time to astronomy. Although hardly mentioned in these two volumes, the Dining Club, at which the



Another world—the Martian landscape as seen by Viking 1. The picture is taken from the new second edition of *Discovering Astronomy* by R. Robert Robbins and William H. Jefferys, an illustrated introduction intended for both students and non-scientists. Publisher is Wiley, price is \$38.60, £28.10.

society originated, has continued to meet and flourish up to the present day. For many years it served as an informal link between the leading Fellows.

Throughout the nineteenth century, the amateur members played a leading role both in the society and in astronomical research. Much early astrophysics in Britain was carried out by amateurs, in contrast to today when the expense and complexity of astrophysical equipment demands the involvement of professionals. A hundred years ago, professional astronomers tended to be trained in mathematics and mainly concerned with classical astronomy. The generally non-mathematical nature of early astrophysics appealed to amateur astronomers, and they were responsible for advances in spectroscopy, photography and the development of new types of telescope.

Something else that distinguished the first hundred years of the society's life was the high level of controversy. In 1879, for example *Nature* commented:

In the interests of British science we have refrained now for some time from referring to the evil days which have fallen upon one of the most reputable of our learned societies. . . . Surely the Fellows of the Royal Astronomical Society of London are strong enough to remedy such a state of things as this.

Since the editor of *Nature* was a Fellow of the society, he spoke with inside knowledge. Matters of contention arose throughout the period (1820–1920) covered by the first volume. One of the last actions recorded is the election of the first female Fellows in 1916 after years of sporadic debate.

The second volume covers the period 1920–1980. It begins with an intriguing episode in the history of British science—the inclusion of geophysics in astronomy. In most countries, geophysics was naturally associated with geology. In Britain, the adherence of some of the leading early geophysicists to the RAS, and the lack of interest of the Geological Society (not mentioned in this volume), led to geophysics being organized by the astronomers. That has led to some tension over the years (again somewhat underplayed here), but so have the other specializations that have blossomed within the society. The

fact that new activities, such as radio-astronomy, have remained within the fold is a tribute to the flexibility of the RAS.

Until well after the Second World War, the running of meetings varied little for decades. The boom in membership during the 1960s led to a number of changes, but one thing that did not change was the number of amateurs who joined. Their influence, however, was clearly on the decline. The last amateur to be elected president held office 30 years ago. As the introduction to the second volume notes: “by 1980 it had become very rare for someone other than a professional to speak at a meeting, publish a paper or be a member of Council”.

Volume 1 of this history first appeared in 1923, three years after the end-date of the history. Volume 2 appeared late last year, seven years after the 1980 end-date, which was itself an extension of the originally intended 1970 deadline. Although it has been revised for publication, some of the material in this volume was therefore written a long time ago. Nevertheless, the two volumes complement each other well, and provide a valuable picture of how a leading British society has developed during its long career. □

Jack Meadows, formerly Head of the Department of Astronomy and History of Science at the University of Leicester, is a Professor in the Department of Library and Information Studies, Loughborough University of Technology, Loughborough LE11 3TU, UK.

New journals review

On 29 September *Nature* will publish the eighth annual review supplement devoted to science journals. Criteria for inclusion of a journal in the 1988 issue are that:

- (i) the first number appeared, or the journal was retitled, between June 1986 and May 1987 (the second cut-off date allows at least three issues of a journal to have been published, the minimum number on which a reasonable judgement can be based);
- (ii) it is published at least three times a year;
- (iii) the main language used is English.

Publishers and learned societies are invited to send four different issues of each suitable periodical, including the first and most recent numbers (if from outside the United Kingdom, by air mail) to: The Review Editor, *Nature*, 4 Little Essex Street, London WC2R 3LF, England.

Genetic packages

Rodger Staden

Sequence Analysis in Molecular Biology: Treasure Trove or Trivial Pursuit? By Gunnar von Heijne. *Academic*: 1987. Pp.188. \$27.95, £17.50.

THE methods used to determine DNA sequences are very much faster than the experiments employed to interpret them. Consequently we have to use computers to analyse the sequences and the proteins they encode, and there is a growing number of programs and techniques available for the job. Most users of these programs are not mathematicians or statisticians, but are biologists who are determining a sequence as part of some longer-term scientific project.

The way in which individuals use computing power to analyse their sequences varies enormously. There are still those who are intimidated by computers, while others positively enjoy sitting at a keyboard and trying out all the available forms of analysis. Many people are limited in what they can do, either by being committed to a particular package of programs or simply because they are unaware of other methods that are available.

Publications describing sequence-analysis software range from the purely mathematical or algorithmic, through more general outlines of what can be done, to thinly disguised advertisements. Von Heijne's book bridges the gaps between these types of publication. It can be read with benefit by all classes of user and also those contemplating entering the field, because it concentrates on conveying the main ideas behind the methods, makes comparisons between them and discusses their reliability.

Most aspects of the subject are covered, ranging from programs to aid laboratory experiments, through DNA-sequence interpretation and prediction of protein secondary structure, to methods of sequence comparison and alignment. The book contains many figures taken from the research literature and is written in a fairly consistently chatty style that makes for easy reading. It is a good introduction, and many readers may subsequently feel more interested in reading the original papers.

The stated take home message is “Don't expect your computer to tell you the truth”. I believe that the book will indeed encourage readers to think more clearly about their use of computers for analysing sequences; to ask themselves what questions they really want to answer; and to be more critical of the results they obtain.

Rodger Staden is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

In the field of the science of string

C.H. Llewellyn Smith

Gauge Fields and Strings. By A.M. Polyakov. Harwood: 1987. Pp.301. Hbk \$48, £32; pbk \$18, £12.

THIS is a unique book. It can and should be read for profit and pleasure by all advanced students of field theory, and by professional mathematical and theoretical physicists.

Polyakov is an outstanding theoretician who has made many seminal contributions to particle physics and statistical physics. He describes the origin of his book in the preface:

For many years I have been keeping notes on different topics in physics — a kind of scientific diary. They contain occasional new results and mostly derivations of known things, done in a way that seemed nice to me. . . . This book has arisen from these notes, or better to say, from the part of them devoted to field theory. . . . In many cases I discuss things that have never been completely understood. I do this in the hope that the approach I suggest, although imperfect, will stimulate deeper penetration into the subject.

As this passage suggests, the book conveys the author's personal approach to subjects on which his own work, and that of other members of the 'school' of Russian physicists to which he belongs, has had a profound influence. It starts with a discussion of the analogy between statistical mechanics and quantum field theory, which is then exploited throughout the book as Polyakov moves back and forth between statistical systems, such as the Ising model, and continuum field theories in a way that illuminates both. Much of the material in Chapters 1–8 — asymptotic freedom, the strong coupling expansion, instantons (which were discovered by Polyakov and co-workers), the topology of gauge fields and the large N expansion — is now included in advanced textbooks on field theory, but Polyakov's treatment is more wide ranging and goes on in content, and also in style and approach, to the frontiers of research.

The climax of the book is the 100-page Chapter 9, "Quantum Strings and Random Surfaces", which is followed by a final chapter entitled "Attempt at Synthesis". Polyakov approaches string theory from the so far incomplete attempt to formulate quantum chromodynamics, the theory of the sub-nuclear inter-quark force, entirely in terms of Wilson's loop variables, which was developed largely by the Russian school. The treatment is centred on Polyakov's own contributions to the definition and calculation of functional integrals for strings which has

been the basis of much recent progress. It is only latterly that gravity enters and the currently fashionable idea that the underlying 'unified theory of everything' is based on strings moving in nine space dimensions, six of which have curled up to minute sizes, is not discussed until Chapter 10. This ordering of the material is an important reminder that even if string theory is not relevant for fundamental unification, it is likely to find applications in quantum chromodynamics and/or in the theory of random surfaces.

Polyakov's book is a supplement to, not a substitute for, the standard books on advanced field theory and will be inaccessible to readers who have not already acquired considerable sophistication in the subject. It is not necessary to be a mathematical field theorist to benefit

from it, however, because Polyakov does not use the language of topology and differential geometry and continually resorts to physical arguments, explanations and analogies. As a relatively pragmatic user rather than producer of field theory, I found this volume a source of enlightenment on many of the developments that have taken place during the remarkable renaissance of field theory in the past 15 years which I have only partially absorbed by osmosis from my colleagues. I have learned a great deal from a first quick reading, and I look forward to studying the book in more detail and referring to it in the future. □

C.H. Llewellyn Smith is Chairman of the Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

Simian society

Carel P. van Schaik

Primate Social Systems. By Robin I.M. Dunbar. Croom Helm/Cornell University Press: 1988. Pp. 373. Hbk £35, \$49.50. pbk £17.50, \$24.95.

WHY is there so much variation in the social systems of the 180 or so species of primate? In the early 1960s, the findings of field workers seemed to indicate clear-cut associations between the habitat, diet and activity period of a species and selected features of its society. The flurry of activity that followed, however, produced many exceptions but no prospects for a new framework.

During the 1970s most field studies concentrated either on ecology or on aspects of social behaviour, such as infanticide and mating strategies. The new concepts of sociobiology made it clear that only the social strategies of individuals are directly affected by ecological conditions. Hence, the links between ecology and the social system are indirect, and therefore complicated.

Robin Dunbar's *Primate Social Systems* is the first book to explore systematically the effects of ecological and demographic factors on the social strategies of individuals and the structure of their societies. In the opening chapters he provides a pithy introduction to the main concepts of evolution and demography that enable us to frame and answer questions about functional aspects of primate social behaviour. After treating the ecology of grouping patterns, he discusses how animals can best obtain their mates, rear their young and use their conspecifics as allies. In the final chapters, he extrapolates from the individual level to that of the society as a whole by developing quantitative models and examining their dynamics.

The great strength of the book is in the detailed analyses of individual variation in social strategies. Dunbar is at his best when he discusses the feedback between demography and social behaviour. We learn that the balance of give and take in a relationship is negotiable, and that patterns in alliance formation depend on the availability of suitable partners. Both of these change in a predictable way as demography or social context changes. Dunbar stresses that such demographic factors can explain much of the inter-specific variation in the structure of primate societies.

Where Dunbar treats ecological questions, his grasp is less firm. The tests that he proposes are not always adequately derived or executed. For instance, to test the hypothesis that communal defence of food sources is the main benefit of group living in primates, he erroneously predicts that groups suffer most from between-group food competition at low population densities. In some tests he does not eliminate possible confounding effects, or he inadvisedly fits curves by eye.

We are only just beginning to understand the intricate links between external conditions and the social behaviour of primates. Some of Dunbar's conclusions may be wrong, but his reasoning is always explicit; and as the ecologist Robert MacArthur once said, it is far more important to be interesting than it is to be right. □

Carel P. van Schaik is a Senior Research Fellow with the Royal Netherlands Academy of Arts and Sciences (KNAW), Laboratory of Comparative Physiology, University of Utrecht, 3572 LA Utrecht, the Netherlands.

More primatology

With the appearance of Vols 2B (*Behavior, Cognition, and Motivation*) and 4 (*Neurosciences*) Alan R. Liss has recently completed publication of the mammoth multi-author work *Comparative Primate Biology*. The books will be reviewed in a forthcoming issue of *Nature*.

Small-scale analog applications of high-transition-temperature superconductors

John Clarke

Department of Physics, University of California, Materials and Chemical Sciences Division, Lawrence Berkeley Laboratory, Berkeley, California 94720, USA

The first practical applications of high temperature superconductors are likely to be small-scale electronic devices such as detectors or analog processors. Several types of superconducting quantum interference devices SQUIDS have already been operated successfully as detectors of magnetic fields. But although operation at higher temperatures, say 77 K, will be more convenient, the sensitivity of all detectors will inevitably be degraded.

GREAT furore surrounds the discovery of superconductivity^{1,2} in ceramic oxides such as $\text{Y}_1\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) with transition temperatures (T_c) approaching 100 K. Apart from the continuing debate over the physical origin of the superconducting phase, the excitement of course stems largely from potential applications of superconductors operating, for example, at the boiling point of liquid N_2 (77 K). Traditional superconductors such as Nb, NbTi or Pb operating at or near the boiling point of liquid ^4He (4.2 K) are used in two broad areas: large-scale 'machines', usually involving wire, and small-scale electronics, usually involving thin films. The reason usually quoted for the advantage of 77 K operation is that liquid N_2 is much cheaper than liquid ^4He . Certainly for larger systems, however, the cost of the helium is not a major fraction of the total cost, although the fractional cost savings for, say, an array of geophysical magnetometers might be larger if one were to replace liquid ^4He with liquid N_2 . But the real advantages lie more in the simplification of the cryostat, in the ready availability of refrigerators capable of cooling to 77 K, and in the fact that the latent heat of vaporization of liquid N_2 is about 60 times higher than that of liquid ^4He . As a result, a cryostat filled with ^4He that is capable of operating unattended for a week, for example, in a remote desert area, would be able to operate for roughly a year when filled with N_2 . Even if one were content to operate devices immersed in liquid Ne at 28 K, the running time would be 40 times greater than for liquid ^4He , although the cost of a litre of Ne is considerably greater than that of He.

The major large-scale applications (for reviews see refs 3 and 4) involve high-field magnets used, for example, in magnetic resonance imaging of the human body, in the acceleration of charged particles in machines such as the Tevatron (Fermi National Accelerator Laboratory), in the separation of magnetic ore from a slurry, and in laboratory experiments—cyclotron resonance or magnetoresistance, for instance. Other proposed applications include generators, transformers and transmission lines for electrical power, energy storage, magnets for fusion or magnetohydrodynamics, and motors for ship propulsion. The field of superconducting electronics (for reviews, see refs 4–8, 48, 49) started in earnest in the early 1960s, and a small number of devices have now emerged that are better than the competing technologies and are used very successfully in an intriguing number of applications, ranging from radio astronomy to the measurement of tiny magnetic fields emanating from the human brain.

Oversimplifying somewhat, we can conveniently divide the field of small-scale electronics into three broad categories, indicated schematically in Fig. 1. First, the input signal and its accompanying noise are fed into an analog detector: this signal could be anything from the 100-GHz signal from a radiotele-

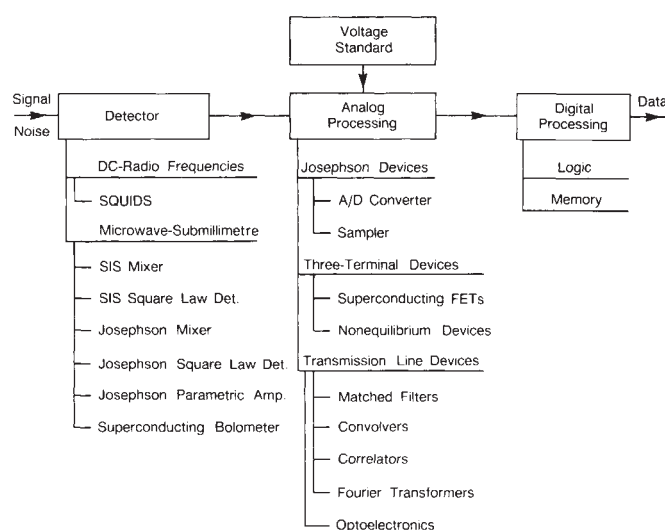


Fig. 1 Superconducting electronics divided into three broad categories: for the detection of an incoming signal, for the analog processing of the amplified signal, and for the subsequent digital processing. Representative examples of each category are listed.

scope to the 1-Hz signal from a human brain. The second stage of processing generally also involves analog devices: for example, the bandwidth of the amplified signal might be narrowed to reduce the noise. Finally, most systems today involve digital processing to produce the data of interest. Substantial effort has been expended to develop logic and memory devices based on Josephson tunnel junctions⁹ for superconducting computers¹⁰, and a cross-sectional module¹¹ of a computer has been successfully demonstrated. At present, however, the only substantial effort on 4.2 K devices is in Japan. It seems unlikely that a large computer based on this technology will be built in the near future, but the prospects for extremely fast micro-processors are excellent.

Here I will confine my attention to the two classes of analog devices, beginning with a general comment. The important superconducting detectors are successful because they have lower noise than competing devices, and this advantage arises, at least in part, because they operate at low temperatures, say 4.2 K. Operating the same devices at a higher temperature, say 77 K, inevitably means that the noise will be increased. Thus, one can never expect the performance of superconducting detectors at 77 K to match that at 4.2 K, although a given device may still be useful if it is the quietest available at 77 K. On the other

Josephson tunnelling

The central feature of superconductivity is the Cooper pair, that is, two electrons with opposite spin and, in the absence of applied currents or magnetic fields, equal and opposite momenta, so that the net spin and momentum of a pair are zero. All pairs are condensed into a single macroscopic quantum state with a phase ϕ that, in the absence of currents or fields, has the same value throughout the superconducting body. Thus, Cooper pairs exhibit long-range phase coherence. In a Josephson tunnel junction a region of the superconductor is weakened, prototypically by a tunnel barrier, as shown in Fig. 2a. When the current I is increased from zero, Cooper pairs tunnel quantum mechanically through the barrier, establishing a phase difference δ between the phases of the two superconductors according to the Josephson current-phase relation $I = I_0 \sin \delta$. The voltage across the junction is zero for $I < I_0$, the critical current of the junction: we refer to this regime as the dc Josephson effect. When the current through the junction is raised to a value $I > I_0$, a voltage V is developed across the junction. The phase difference δ increases with time t as $d\delta/dt = 2eV/\hbar$, where e is the electronic charge and $2\pi\hbar = h$ is Planck's constant. We refer to this phenomenon as the a.c. Josephson effect.

A practical Josephson tunnel junction using conventional superconductors is shown in Fig. 2b. A thin film of a superconducting material, for example niobium, is deposited on an insulating substrate and an oxide layer, typically 20–30 Å thick, is grown on it. This barrier may be formed in several ways, for example, by thermal oxidation of the superconductor or by the deposition and subsequent oxidation of aluminium¹². A second superconducting film is deposited to cross the first film, thus forming the tunnel junction. Modern photolithographic techniques are routinely used to produce junctions 1–2 μm across, the area of the junction often being defined by a window in a thick insulating layer such as SiO₂. Electron-beam lithography can produce junctions with areas as small as 10⁻² μm² or even less. A very high degree of reproducibility and reliability has been achieved in the fabrication of junctions from niobium.

When the junction is cooled well below the transition temperature of the superconductors, one obtains the current-voltage (I - V) characteristic illustrated in Fig. 2c. As the current is increased from zero, initially no voltage appears: this is the d.c. Josephson effect. When I exceeds I_0 , however, the voltage switches abruptly to $\sim 2\Delta/e$, roughly 3 mV for Nb (Δ is the energy gap of the superconductors). As we increase and subsequently

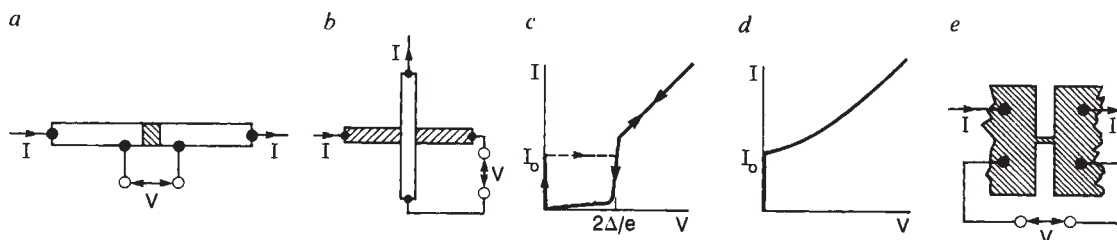


Fig. 2 a, Schematic representation of Josephson tunnel junction showing two superconductors separated by an insulating barrier. Passage of a current through the junction involves the quantum mechanical tunnelling of Cooper pairs through the barrier. When the current exceeds the critical current I_0 of the junction, a voltage V appears. b, Practical cross-strip tunnel junction. A thin-film of (conventional) superconductor is oxidized (shading), and a second film deposited over the oxide layer. When the junction is cooled below the superconducting transition temperature, the current-voltage (I - V) characteristic shown in c is obtained as the current is increased from zero to a maximum value and then reduced to zero again. The hysteresis evident in c can be eliminated by means of a thin film of normal metal connecting the two superconducting films to provide a shunt conductance in parallel with the tunnel junction: the resulting non-hysteretic I - V characteristic is shown in d. e, The Anderson-Dayem bridge. The thin film is necked down to produce a very narrow channel connecting the two superconductors.

decrease the current, a hysteretic I - V characteristic is traced out. This hysteretic behaviour forms the basis of computer memory and logic devices, but must be eliminated for certain other devices, notably SQUIDs. The non-hysteretic characteristic shown in Fig. 2d is obtained by shunting the junction with a thin normal film of sufficiently low resistance.

Unfortunately, it is not trivial to extend the existing Josephson junction technology to a material such as YBCO. One problem is the difficulty of maintaining the integrity of a thin insulating barrier during the processing of the counter-electrode at (say) 650 °C. A second problem is the fact that the surface layers of YBCO appear to be non-superconducting. Substantial efforts are already underway to attempt to circumvent these problems.

Finally, Fig. 2e shows a quite different type of junction, the Anderson-Dayem bridge¹³. The bridge is itself superconducting, but with a sufficiently small cross-section that it can sustain only a tiny supercurrent. This structure exhibits Josephson-like behaviour with an I - V characteristic something like that in Fig. 2d, provided the length of the bridge is no greater than the coherence length ξ , which is, roughly speaking, the spatial extent of a Cooper pair. In a low temperature superconductor such as Nb, ξ is ~ 100 nm, a dimension that can be achieved with modern lithographic techniques. In YBCO, on the other hand, ξ is estimated to be rather less than, or somewhat greater than, 1 nm, depending on the crystalline orientation, and it has (so far) been impracticable to produce structures of this size.

hand, when one considers analog (or indeed digital) processing devices, the increase in noise does not necessarily degrade the performance.

I will concentrate on a few of the more successful, or potentially important, analog devices, focusing in particular on SQUID—superconducting quantum interference devices—which are based on the phenomenon of Josephson tunnelling (see panel above). Several types of SQUID have already been fabricated from high-transition temperature superconductors. Over the past two decades SQUIDs (see Fig. 3) at liquid ⁴He temperatures have been put to an amazingly broad range of applications, including: (1) laboratory-based instrumentation, for example, voltmeters with sensitivities as high as 10⁻¹⁵ V, radiofrequency amplifiers for detecting magnetic resonance, and susceptometers; (2) the detection of magnetic signals from human subjects, for example, spontaneous or evoked brain activity; (3) geophysics, for example, magnetotellurics, electromagnetic sounding, palaeomagnetism, tectonomagnetism

and internal ocean waves; (4) 'exotic instruments' such as gravity wave detectors and monopole detectors.

My selection of devices is not meant to be exhaustive or to imply that none of the other devices will ever be important, but illustrates the wide variety of superconducting electronics. To simplify the discussion, I have assumed that the transition temperature remains ~ 100 K and that devices are operated at 77 K. Of course, discovery of 'room-temperature superconductivity' ($T_c > 400$ K for 300 K operation) would change everything.

Analog signal processing

One important class of device is for analog processing of high-frequency (say, gigahertz) signals. One example is the improvement of the signal-to-noise ratio in radar receivers or communications receivers by optimal (or matched) filtering, that is, the use of a filter with a frequency response tailored to a specific signal shape. Other examples are the performing of a Fourier transform of a signal to analyse its spectral content, and the

Fig. 3 *a*, The direct current (d.c.) SQUID: two Josephson tunnel junctions (x) connected in parallel on a superconducting loop of inductance L . Each junction is shunted with its self-capacitance C and a resistor R to eliminate hysteresis in the I - V characteristic (as in Fig. 2*d*). The SQUID is biased with a current I to produce a voltage V . As the magnetic flux Φ is changed, the voltage oscillates with a period of one flux quantum, Φ_0 . *b*, The radiofrequency (r.f.) SQUID: a single Josephson junction incorporated into a superconducting loop with inductance L . The loop is inductively coupled to the inductor of a resonant circuit excited at its resonant frequency (typically 20 MHz) with a radiofrequency current. As the flux Φ in the loop is changed, the amplitude of the r.f. voltage V_{rf} is modulated with a period Φ_0 .

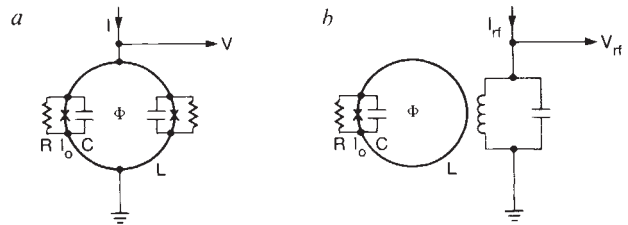
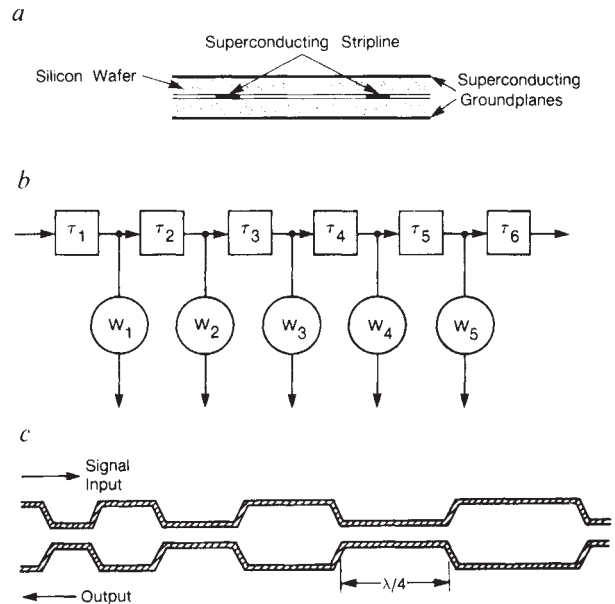


Fig. 4 Analog processing with a superconducting transmission line: *a*, cross-sectional view of a superconducting stripline. The stripline is deposited on the top side of the lower silicon wafer which has a superconducting groundplane on its underside. The upper wafer has a second groundplane deposited on its top side. *b*, The transmission line is used to store a signal propagating (say) from left to right. A series of 'taps' along the line detects the signal as it propagates, thus sampling it at time intervals $\tau_1, \tau_2, \tau_3, \dots$. Each sample is multiplied by a weighting function W_i , and the outputs combined coherently as appropriate for the application. *c*, Filter. The input line is coupled to the output line by taps of progressively increasing length: each tap couples strongly to signals with a quarter-wavelength equal to its length. (Redrawn from refs 14 and 15.)



cross-correlation of two pulses to determine the similarity of their shape as a function of their separation in time. One method of performing these functions involves storing the signal(s) on a delay line. To obtain a sufficiently long delay time on an electromagnetic stripline without making it unreasonably long, one must reduce the width of the line. Unfortunately, the reduced size tends to result in unacceptably high signal losses for lines made of normal metal.

The use of superconducting films overcomes the loss problem¹⁴⁻¹⁶. The stripline is fabricated from niobium deposited on two 50-nm diameter silicon wafers in the cross-sectional geometry indicated in Fig. 4*a*. The line is in a meander or spiral pattern, and is typically ~ 1 m long. The delay line is 'tapped' at regular intervals so that the signal(s) on the line can be sampled at successive time-intervals $\tau_1, \tau_2, \tau_3, \dots$ as indicated in Fig. 4*b*. Each sample can be multiplied by a weight W_i before being coherently combined with the other samples, as appropriate to the application. For example, one kind of filter consists of a stripline carrying the signal coupled to a second parallel line, on which the output propagates in the reverse direction, as shown in Fig. 4*c*. Each 'tap' consists of a region of length l over which the two lines are close together: when this length is equal to $\lambda/4$, where λ is the wavelength of the signal, the coupling is enhanced. Thus, by making a series of taps of the required lengths, one couples only selected wavelengths to the output.

Several analog processors have been operated using niobium films at 4.2 K, and filtering, correlation and Fourier transformation demonstrated¹⁴⁻¹⁶. It is predicted that bandwidths as large as 10 GHz and computational rates equivalent to 10^{12} arithmetic operations per second should be possible. Although this tech-

nique is still in its infancy, it holds considerable promise for future development. With regard to YBCO films, a major concern is whether the signal losses at 77 K will be sufficiently low for this application. Initial measurements of these losses at 4.2 K indicate that they are much higher than in Nb (M. R. Beasley and R. S. Withers, personal communication), but substantial reductions are to be expected as the quality of the films, particularly of the surface, is improved. The advent of high- T_c superconductivity may speed the development of these processors, not least because one could combine the benefits of a low loss-line and fast semiconductor devices, some of which perform very well at 77 K.

A second kind of device that has been operated successfully as an instrument at 4.2 K is the Josephson sampler^{17,18}. Sampling techniques are commonly used to determine the waveforms of very high-speed repetitive signals: the amplitude of the signal is 'sampled' for a very small time interval stepped successively through the pulse. When these samples are displayed as a function of time, for example on an oscilloscope screen, a picture of the waveform is reconstituted. The advantage of the technique lies in the very fast pulses that can be generated.

At the heart of the Josephson sampler is a sampling gate, actually a SQUID, which switches to a new state only if the input current exceeds a certain value, I_g . The signal current $I_s(t)$ to be analysed is triggered at time $t=0$, and delayed by an adjustable time t_0 . At this instant, a current pulse of amplitude I_p and a variable bias current I_b are added to $I_s(t_0)$. If the sum of these three currents exceeds I_g , the gate is triggered. The value of I_b is controlled by a feedback circuit until it is just large enough to trigger the gate; then $I_s(t_0) = I_g - I_p - I_b$. By varying t_0 , one can thus trace out the waveform of the input

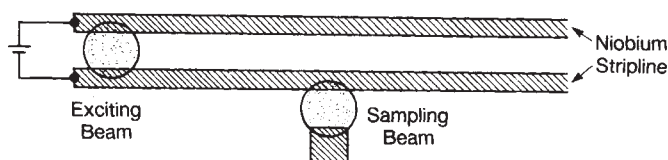


Fig. 5 Optoelectronic sampling device consisting of two Nb strips deposited on a silicon layer. The exciting laser spot generates photocarriers in the substrate which momentarily short together the ends of the line. The resulting pulse propagates down the line and is subsequently sampled by a second laser spot that shorts one side of the stripline to an electrode. (Redrawn from ref. 20.)

signal. Sampling resolution times as low as 2 ps have been achieved¹⁹, about an order of magnitude better than the fastest conventional sampling oscilloscopes. The instrument would be even more useful if it were operated at 77 K because of the relative ease of connecting it to signal sources at room temperature.

More recently, an even higher sampling resolution has been obtained²⁰ by combining optoelectronic techniques with a superconducting transmission line. The line consists of two parallel strips of niobium, a few micrometres wide, deposited on a silicon-on-sapphire substrate (Fig. 5). A voltage is applied to the two strips. To generate a signal, a very short (80 fs) optical laser pulse is focused on one end of the line, generating photoexcited carriers that momentarily short the lines together. The pulse so generated propagates down the line, and is sampled by means of a second laser pulse that momentarily shorts one of the niobium strips to a third sampling electrode. The waveform propagating down the line is reconstituted by varying the time between the initiating and sampling pulses. Pulses with a 0.9 ps full-width at half-maximum have been generated and detected with this technique. The use of YBCO films in this application is obviously very appealing. Potential problems are the requirement of low signal-propagation loss and the difficulty of depositing YBCO films directly on to a semiconducting substrate.

Another potentially important device is the fast analog-to-digital (A-to-D) converter, based on Josephson junctions, that has been developed with Nb- and Pb-technologies^{21,22}. This system might interface the output from the analog processor to the digital processor (Fig. 1). One type of A-to-D converter is based on the dc SQUID (Fig. 3a), which converts a monotonically varying magnetic field into a periodic voltage, thus digitizing the input signal. Systems have been built and tested that are capable of digitizing 6 bits at 100 MHz (ref. 21), 4 bits at 280 MHz (ref. 22), and 3 bits at 500 MHz (ref. 22). It is projected²¹ that these performances can be improved to 6 bits at 300 MHz and 4 bits at 1 GHz. This predicted performance is comparable with, or slightly better than, that of conventional A-to-D converters. Existing superconducting A-to-D converters rely on Josephson tunnel junctions, and, in principle, a similar performance should be possible with high- T_c superconductors, provided a reliable technology for Josephson tunnel junctions can be developed. A fast A-to-D converter operating at 77 K would be attractive because of the possibility of integrating it with semiconductors at the same temperature.

Finally, a well established application of Josephson tunnelling is not actually a processing device, but a very accurate voltage standard that can be used for calibration purposes. The Josephson standard volt was one of the earliest applications of Josephson tunnelling²³, and is the means by which most national standards laboratories maintain the volt. The principle is simple: when a high-capacitance Josephson tunnel junction is irradiated with microwaves of frequency f , constant-voltage current steps are induced at voltages $V = nhf/2e$ (see Fig. 6), where n is an integer, h is Planck's constant and e is the electronic charge. For example, a frequency of 96 GHz corresponds to about 200 μ V for $n = 1$. To facilitate comparison with other voltage

standards, the modern Josephson standard, developed at NBS Boulder, PTB Berlin, and ETL Japan, consists of 1,500 to 2,000 junctions in series to produce a standard of ~ 1 V (ref. 24). As frequency can readily be determined to extremely high accuracy, Josephson junctions provide a straightforward means of producing an accurate volt in any laboratory. Suitable arrays are available commercially. Of course, the procedure would be simplified if the device were to operate at 77 K. Again, there is a pressing need for a Josephson tunnel junction; non-hysteretic junctions could possibly be used, but their operation in an array would not be easy.

Far-infrared detectors

The input end of Fig. 1 lists some of the devices that have been investigated as sensitive detectors of far-infrared electromagnetic radiation over the past two decades. The result has been one of the major success stories of superconducting electronics, the SIS (superconductor-insulator-superconductor) tunnel junction mixer (for reviews, see refs 25 and 26). Mixers of this kind are installed as receivers on a number of radio telescopes²⁷, working mostly at frequencies around 100 GHz, and are typically used to observe molecular lines.

A mixer combines a weak signal at frequency f_s with the output from a local oscillator at frequency f_{LO} to produce sum and difference frequencies $f_s \pm f_{LO}$. The difference frequency $f_s - f_{LO}$ is subsequently amplified by an intermediate frequency amplifier. The mixing process requires a sharp nonlinearity in the current-voltage characteristic, which in the case of the SIS mixer is found at the sharp rise in the current at a voltage $2\Delta/e$ (see Fig. 2c). (In the SIS mixer, the Josephson current is suppressed with a magnetic field.) Carefully implemented, SIS mixers have useful levels of gain, and at 36 GHz have exhibited noise levels within a factor of 2 of the quantum limit²⁸. At frequencies up to several hundred gigahertz this mixer is the most sensitive detector available, and its continued success when operated at liquid ^4He temperatures seems assured.

The most obvious problem in fabricating an SIS mixer from YBCO is the need for a tunnel junction with an exceedingly sharp rise in current at $V = 2\Delta/e$ (see Fig. 2c), combined with very low current leakage at lower voltages. Given the difficulties with making any kind of high- T_c tunnel junction, suitable junctions are unlikely to be available soon. Existing theories for the SIS mixer are for low-temperature operation, and there appear to be no detailed predictions for the performance of an ideal junction at 77 K. Thus, beyond the general statement that the noise is likely to increase as the temperature is raised, it is difficult to assess the potential usefulness of an SIS mixer at high temperatures, even if one can be fabricated.

A quite different type of far-infrared detector is the superconducting bolometer²⁹, which consists of an absorbing substrate supporting a thin film at the midpoint of its superconducting resistive transition. The substrate has weak thermal coupling to a heat bath, and the strongly temperature-dependent resistance of the superconducting film is used to measure the temperature rise resulting from the absorption of radiation. These broadband detectors have been operated successfully at liquid- ^4He temperatures, but are not as sensitive as the best semiconducting bolometers. Superconducting bolometers will be less sensitive at 77 K than at 4.2 K, but might be competitive with other detectors at that temperature.

SQUIDS

The most widely used superconducting device, and one that has been commercially available for many years, is the SQUID (for reviews, see refs 30 and 31). The two varieties, direct current (d.c.) and radiofrequency (r.f.), are shown in Fig. 3. Of the two, the d.c. SQUID is the more highly developed and the more sensitive, but the r.f. SQUID is more widely used, largely because it has been commercially available for almost two decades. When the d.c. SQUID is biased with a constant current I , the voltage

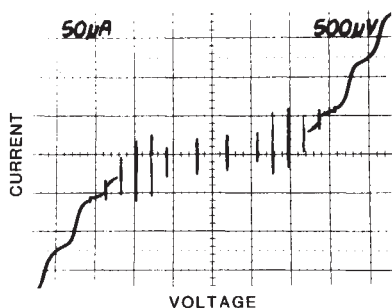


Fig. 6 Current-voltage characteristic of Josephson tunnel junction in the presence of $f = 96$ GHz microwaves. Current steps are produced at voltages of precisely $\pm n\hbar/2e$, where n is an integer. In the absence of microwaves, the I - V characteristic is similar to that in Fig. 2c (from ref. 24).

V across it oscillates as a function of the applied flux Φ with period $\Phi_0 = h/2e \approx 2 \times 10^{-15}$ Wb. Using niobium-based thin-film technology several groups have produced relatively sophisticated devices, usually involving the thin-film 'washer' design³². An insulating layer is deposited over the SQUID, followed by a multi-turn spiral coil of superconducting film, through which the input current passes. This coil is coupled to the circuitry appropriate for a given application.

The r.f. SQUID (Fig. 3b) is coupled to a tuned circuit excited by an r.f. current, typically at 20 MHz; the amplitude of the r.f. voltage, V_{rf} , is periodic in Φ . Commercially available r.f. SQUIDS consist of a toroidal body machined from niobium, inside which is a single Josephson junction. The input coil, wound from niobium wire, is placed in the toroidal cavity.

Both types of SQUID are flux-to-voltage transducers, producing an output voltage in response to a magnetic flux Φ . One usually operates them in a feedback circuit that produces an equal and opposite flux to cancel any applied flux, thereby both linearizing the response of the SQUID and enabling one to detect changes in flux much less than Φ_0 .

To assess the impact of higher operating temperatures on SQUIDS, we need to characterize their resolution. The simplest way is in terms of the equivalent flux noise, which represents the smallest change in magnetic flux the SQUID can resolve. For example, the best d.c. SQUIDS in the liquid-⁴He temperature range can resolve a few $\mu\Phi_0 \text{ Hz}^{-1/2}$ at frequencies above the $1/f$ (flicker) noise region where the noise is white. A more useful means of characterizing the resolution is in terms of the flux noise energy per unit bandwidth, $\epsilon = \langle \Phi_N^2(f) \rangle / 2L$, where $\langle \Phi_N^2(f) \rangle$ is the mean-square flux noise per unit bandwidth, and L is the inductance of the SQUID. At 4.2 K, typical state-of-the-art thin film d.c. SQUIDS have noise energies in the range $(1\text{--}5) \times 10^{-32} \text{ J Hz}^{-1}$ (or, $100\text{--}500 \hbar$), although values as low as a few $\hbar$ have been achieved^{33–35}. Typical commercial d.c. SQUIDS, on the other hand, have noise energies around $10^{-30} \text{ J Hz}^{-1}$ ($10,000 \hbar$). Commercially available r.f. SQUIDS are somewhat less sensitive, with noise energies of $\sim 10^{-29} \text{ J Hz}^{-1}$ ($100,000 \hbar$).

The d.c. SQUID has a well established theory³⁶ for white noise, and we will use it to predict their performance at 77 K; a somewhat similar discussion has been given by Pegrum and Donaldson³⁷. First, we note that for quantum effects to be observed, the root-mean-square flux noise in the SQUID loop, $(k_B T L)^{1/2}$, must be less than $\Phi_0/2$ (where k_B is Boltzmann's constant and T is the temperature). The resulting requirement at 77 K, $L < 1 \text{ nH}$, is actually satisfied by most SQUIDS currently operated at 4.2 K. Second, the coupling energy of each junction, $I_0 \Phi_0 / 2\pi$ (I_0 is the critical current or maximum supercurrent), must be much greater than $k_B T$; this condition leads to $I_0 \gg 3 \mu\text{A}$ at 77 K. If we take as arbitrary, but reasonable, values $L = 0.2 \text{ nH}$ and $I_0 = 20 \mu\text{A}$, we find $\beta = 2LI_0/\Phi_0 = 4$, a value not too far from the optimal value $\beta = 1$. Finally, we assume that we connect

a shunt across the junction with resistance R , just small enough to eliminate the hysteresis on the current-voltage characteristic: say $\beta_c = 2\pi I_c C R^2 / \Phi_0 \approx 0.5$, where C is the self-capacitance of the junction. When these various conditions are met, the noise energy is given by³⁶

$$\epsilon = \alpha k_B T (LC)^{1/2} \quad (1)$$

where α is a number varying slowly with the parameters, with a typical value of 30. Equation (1) has two obvious implications: smaller is better, and cooler is better. We can estimate the performance expected for SQUIDS based on thin-film tunnel junctions at 77 K by scaling equation (1), assuming L and C remain fixed, to find ϵ in the range $2,000\text{--}10,000 \hbar$. Thus, although the performance degrades by a factor of about 20 over that at 4.2 K, it is nonetheless comparable with the d.c. SQUIDS currently commercially available.

Many applications of SQUIDS are at low frequencies, around 1 Hz, and $1/f$ noise is therefore of paramount importance. There are at least two known sources of $1/f$ noise in SQUIDS at liquid ⁴He temperatures. The first arises from the trapping and detraping of electrons³⁸ in the barrier of the tunnel junction, a process producing $1/f$ noise in I_0 . The noise is much reduced by the use of high-quality, low-leakage tunnel junctions^{39,40}, and can be further reduced by appropriate flux modulation schemes^{40,41}. The second source is less well understood, but appears to be related to the motion of magnetic flux trapped in the body of the SQUID (ref. 41). This 'flux noise' depends strongly on the material properties of the SQUID and can be very low. Unfortunately, both sources of $1/f$ noise increase rapidly with temperature, so that the level of noise at 77 K may be high enough to be of concern.

The r.f. SQUID also has a well developed noise theory³¹. The intrinsic noise energy may be written in the form

$$\epsilon \approx \alpha' \frac{LI_0^2}{\omega_{rf}} \left(\frac{2\pi k_B T}{I_0 \Phi_0} \right)^{4/3} \quad (2)$$

where $\alpha' \sim 1$ and ω_{rf} is the bias frequency. Taking the same parameters as for the d.c. SQUID, $L = 0.2 \text{ nH}$ and $I_0 = 20 \mu\text{A}$, and with $\omega_{rf}/2\pi = 20 \text{ MHz}$, at 77 K we find $\epsilon \approx 10^{-28} \text{ J Hz}^{-1}$. The concerns about the $1/f$ noise are similar to those for the d.c. SQUID.

High- T_c r.f. SQUIDS

Both r.f. and d.c. SQUIDS were produced remarkably quickly after the discovery of superconductivity in YBCO. The first r.f. SQUIDS^{42,43} consisted of a pellet of sintered YBCO placed in the inductor of an LC -resonant circuit. It was found that the polycrystalline material behaved as an r.f. SQUID, or, to be exact, several r.f. SQUIDS. The amplitude of the r.f. voltage was modulated by a flux applied to the YBCO, with multiple periods corresponding to loops with radii $7\text{--}100 \mu\text{m}$ (ref. 43). As the sizes of the YBCO grains were in the range $10\text{--}50 \mu\text{m}$, it appears that each SQUID consisted of a small number of superconducting grains interconnected by weak links at their boundaries. When operated in flux-locked loop, the SQUID at 77 K achieved a noise level of about $0.2 \text{ nTHz}^{-1/2}$ at frequencies around 100 Hz. This performance is comparable with that of a flux-gate magnetometer. Because the SQUID parameters are unknown, it is impossible to compute a noise energy.

Zimmerman and co-workers⁴⁴ produced the r.f. SQUID, shown in Fig. 7, that uses a 'break junction'. A cylindrical pellet of YBCO, with a hole drilled through and a slot cut part way along a radius as indicated, was glued into an aluminium holder, and the whole assembly cooled to 77 K. A taper pin was forced into the slot in the mount, thereby causing the YBCO to break in the region of the cut; when the pin was withdrawn slightly, the YBCO surfaces on the two sides of the crack were brought into contact, producing a junction. This technique allows the YBCO to be broken under liquid helium, and so brings together two uncontaminated surfaces. The r.f. SQUID so formed was

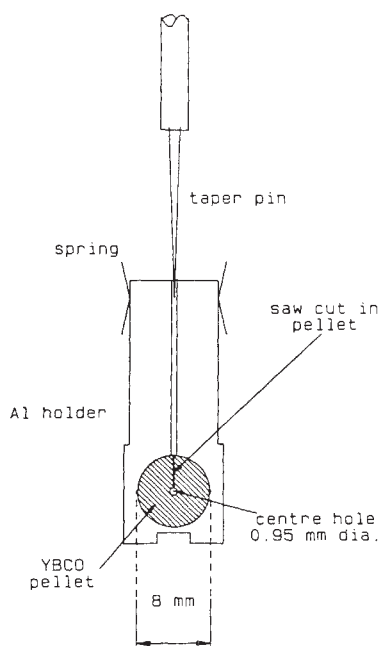


Fig. 7 Radiofrequency SQUID of Zimmerman *et al.*⁴⁴ made from bulk YBCO. A cylinder of YBCO (shaded) with a centre hole is partially cut along a radius. The material is broken by means of a taper pin that forces apart the aluminium holder in which the pellet is glued; the two sides of the saw-cut are subsequently allowed to touch to form a 'break junction'. The SQUID is inductively coupled to a radiofrequency resonant circuit, as in Fig. 3b. The amplitude of the radiofrequency signal is periodic in the flux threading the hole in the pellet (from ref. 44).

coupled to a resonant circuit and operated in the usual way, with an output that oscillated smoothly as a function of magnetic field. The best magnetic flux resolution quoted (at 75 K) is $4.5 \times 10^{-4} \Phi_0 \text{ Hz}^{-1/2}$; with a hole diameter of 0.95 mm, this corresponds to a magnetic field resolution of about $10^{-12} \text{ T Hz}^{-1/2}$. Using the estimated inductance of 0.25 Hz, we find a noise energy of $1.6 \times 10^{-27} \text{ J Hz}^{-1/2}$.

High- T_c d.c. SQUIDS

Iguchi and co-workers⁴⁵ attached leads to a piece of sintered YBCO $\sim 10 \text{ mm} \times 5 \text{ mm} \times 1 \text{ mm}$, and used it as a d.c. SQUID. On cooling the device to 77 K, they found that the critical current was relatively large, and not modulated by an external magnetic field. They then struck the material with a hammer, apparently inducing cracks, and simultaneously reducing the total critical current and creating two or more Josephson-like weak links. Subsequently, they found that the critical current was modulated by a magnetic field applied perpendicularly to the plate, often exhibiting two or more periods. The period of one particular device was about 2 μT , corresponding to a loop area of about $1,000 \mu\text{m}^2$, suggesting that the weak links were formed by contacts between grains in a manner rather similar to that observed for the r.f. SQUID. They did not report measurements of the noise but, from the data given in the paper, we can make a rough calculation of the expected sensitivity. We estimate $\partial V / \partial \Phi \approx 1 \mu\text{V} \Phi_0^{-1}$ and a resistance per junction (assuming two junctions in parallel) of $R \approx 15 \Omega$. The optimum voltage noise³⁶ is $V_N \approx (16 k_B T R)^{1/2} \approx 0.5 \text{ nV Hz}^{-1/2}$ at 77 K. These values lead to a predicted flux noise $V_N (\partial V / \partial \Phi)^{-1}$ of $0.5 \text{ m}\Phi_0 \text{ Hz}^{-1/2}$ and a corresponding magnetic flux sensitivity of $1 \text{ nT Hz}^{-1/2}$.

The first high- T_c thin-film SQUID was developed by Koch *et al.*⁴⁶, and is shown schematically in Fig. 8. YBCO films were evaporated from three sources on to MgO or sapphire substrates held at 400 °C in an oxygen pressure of 10^{-4} to 10^{-3} torr, and subsequently annealed at 900 °C in oxygen. The film was covered with a gold film 0.5 μm thick, which was patterned to leave

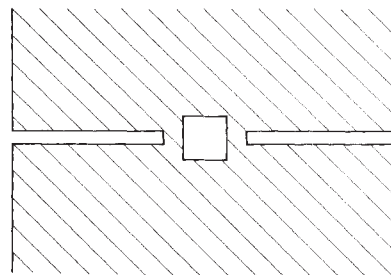


Fig. 8 Direct current SQUID of Koch *et al.* made from thin film of YBCO (shaded area) evaporated on to a sapphire substrate. The two weak links are formed one on each side of the square hole, which is 40 μm square. Each weak link is believed to consist of a grain boundary, or possibly several in series. A current is passed through the two weak links to develop a voltage across them as in Fig. 3a. This voltage is periodic in the flux in the central hole (redrawn from ref. 46.)

covered the portions of the YBCO that were to remain superconducting. The films were ion-implanted with O_2 or As, so that the unprotected regions of YBCO became insulating at low temperatures; the gold mask was removed by ion-milling. The two bridges in Fig. 8 were 17 μm wide, and the loop was $40 \mu\text{m} \times 40 \mu\text{m}$, giving an estimated inductance of 80 pH. The critical current of the device reported varied from about 150 μA at 4.2 K, to about 10 μA at 60 K. It appears likely that the junctions were formed between superconducting grains in the bridges. Modulation of the critical current of the SQUID by an external magnetic field was observed at temperatures up to 68 K; at 40 K the measured flux noise was less than $1 \text{ m}\Phi_0 \text{ Hz}^{-1/2}$ at 100 Hz, limited by noise in the amplifier coupled to the SQUID. The corresponding magnetic field sensitivity (enhanced by a factor of 5 by the 'flux-focusing effect' of the square washer) was $0.25 \text{ nT Hz}^{-1/2}$, and the noise energy $2.5 \times 10^{-26} \text{ J Hz}^{-1}$.

A similar d.c. SQUID has been operated by Nakane *et al.*⁴⁷, who used conventional lithography and an acid etch to produce a $0.2 \text{ mm} \times 1 \text{ mm}$ hole in the YBCO film. The bridges were 0.2 mm wide. These workers reported SQUID behaviour at temperatures up to 72 K, but did not measure the noise. From their results however, we can make a rough estimate of the expected white noise. The published V - Φ plot at 65 K yields $\partial V / \partial \Phi \approx 1 \mu\text{V} \Phi_0^{-1}$. From the measured I - V characteristics at 63 K we estimate a resistance per junction of $R \approx 0.2 \Omega$, which leads to an optimum noise voltage³⁶ of $V_N \approx (16 k_B T R)^{1/2} \approx 5 \times 10^{-11} \text{ V Hz}^{-1/2}$ at 65 K. We thus predict $\Phi_N = V_N (\partial V / \partial \Phi)^{-1}$ to be not less than $5 \times 10^{-5} \Phi_0 \text{ Hz}^{-1/2}$ at frequencies above the $1/f$ noise region.

Comments on high- T_c SQUIDS

We have seen that a variety of high- T_c SQUIDS have already been operated, and doubtless more will be devised in the near future. It is likely that the thin-film devices will be the most widely developed, with configurations similar to those of 4.2 K devices. It should be noted that the performance of the thin-film d.c. SQUIDS described here are substantially below predictions, probably because the junctions were non-ideal: the theory, or course, assumes ideal tunnel junctions. One may hope for rapid progress on the technology of Josephson junctions or weak links in the near future, with a concomitant improvement in the sensitivity of SQUIDS.

To give a perspective on the sensitivities achieved so far, we note that for many applications, for example, instrumentation and geophysics, a noise energy of $10^{-28} \text{ J Hz}^{-1}$ would be adequate, provided it extends down to $\sim 1 \text{ Hz}$. For biomedical applications a noise energy of 10^{-30} J Hz down to 1 Hz is required. The magnetic field sensitivities for the high- T_c devices are not particularly good: a state-of-the-art flux-gate mag-

netometer operating at room temperature may approach 1 pT Hz^{-1/2}. But I emphasize that virtually all practical SQUID magnetometers use a flux transformer, that is, a relatively large superconducting pick-up loop coupled to the input coil. This transformer, which operates down to zero frequency, may improve the magnetic field sensitivity by very substantial factors. A suitable high- T_c flux transformer will be a vital part of the technology of high- T_c SQUIDs.

The likely impact of 77 K operation on the various applications of SQUIDs varies considerably. For example, one could imagine that a sensitive voltmeter able to operate at 77 K, rather than at 4.2 K, would greatly enhance its usefulness. Similarly, geophysical devices that usually operate in remote areas would be much easier to deploy in liquid N₂ (or Ne) cryostats than they are in liquid ⁴He cryostats. On the other hand, the SQUID used in the transducer of a gravity-wave detector should ideally be quantum-limited, and cooled down to millikelvin temperatures, rather than warmed up to 77 K. In the case of biomagnetic sources, currently the largest commercial market for SQUIDs, the cost of the cryogenics is a relatively small fraction of the capital and operating costs, and the advantages of operating at high temperatures with a less sensitive device may not be so obvious. Perhaps a high- T_c flux transformer could be used with a 4.2 K SQUID, thus bringing the pick-up loop(s) closer to the skull of the subject (D. B. Crum, personal communication). But these applications are merely extensions of 4.2 K technology: much more intriguing is the possibility of hitherto unknown applications of SQUIDs made possible by the relative ease of operation at 77 K.

Conclusions

Superconducting electronic devices have been around for a quarter of a century and, as in any technology, many devices have been investigated, found wanting in some respect, and discarded. The only devices presently used in paractical application are SQUIDs, SIS detectors and the standard volt. As we have seen, high- T_c SQUIDs have already been fabricated, but their performance will have to be improved if they are to find significant applications. The most immediate requirement is the development of a reliable thin-film tunnel junction or weak link, not only for SQUIDs but for the standard volt, the sampler,

and for other potentially important devices like the A-to-D converter. The transmission line devices, on the other hand, do not necessarily involve junctions, and we may hope for relatively rapid development, provided the surface losses can be reduced to an acceptable level and the reaction of the films with the substrate reduced sufficiently.

Finally, it is exceedingly important to realize that the new materials have not in themselves given us concepts for new devices. I am optimistic that they will, that the vastly greater numbers of scientists now involved in this field will invent new 'widgets': for example, we can now think seriously of making superconductor-semiconductor hybrid circuits. At the very least, I expect to see more widespread applications of presently known devices, operated at higher temperatures. We could conceivably see exotic applications such as high- T_c devices operating at the ambient temperature of a deep-space probe with no on-board refrigerator. But if small-scale, high- T_c devices are going to have a major impact on our society or economy, I strongly suspect it will be because of devices or applications not yet conceived.

This article was submitted on 18 January 1988, following presentation at Perspectives in the New Superconductivity, Boston, sponsored by *Nature*. I thank M. R. Beasley, F. Bedard, D. B. Crum, R. H. Koch, P. L. Richards, A. D. Smith, T. van Duzer and R. S. Withers for discussions, Drs G. B. Donaldson, C. M. Pegrum, R. H. Koch, I. Iguchi, R. S. Withers and J. E. Zimmerman for preprints of their work, and C. A. Hamilton and J. E. Zimmerman for Figs 6 and 7 respectively. This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy.

Note added in proof: In recent work on their thin-film d.c. SQUIDs, Koch and coworkers (R. H. Koch, personal communication) have obtained a flux noise of about $10^{-5} \Phi_0 \text{ Hz}^{-1/2}$ at a measurement frequency of 1 kHz and a temperature of 40 K. The corresponding noise energy is $\sim 3 \times 10^{-30} \text{ J Hz}^{-1}$. The spectral density of the flux noise scales as $1/f$, giving a noise energy of $\sim 3 \times 10^{-27} \text{ J Hz}^{-1}$ at 1 Hz. Recent measurements⁵⁰ on flux noise in thin-film loops of YBCO have yielded comparable values of $1/f$ noise, and shown that this noise increases rapidly as the temperature is increased from 4.2 K to the transition temperature.

- Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189-193 (1986).
- Wu, M. K. *et al. Phys. Rev. Lett.* **58**, 908-910 (1987).
- Superconducting Machines and Devices* (eds Schwartz, B. B. & Foner, D.) 1-692 (Plenum, New York, 1974).
- Proceedings of the Applied Superconductivity Conf. IEEE Trans. Magn.* MAG-17, no. 1 (1981); MAG-19, no. 2 (1983); MAG-21, no. 3 (1985); MAG-23, no. 2 (1987).
- Superconductor Applications: SQUIDs and Machines* (eds Schwartz, B. B. & Foner, S.) 1-737 (Plenum, New York, 1977).
- IEEE Trans. Electron. Devices* ED-27 No. 10 (1980).
- Proceedings of the International Conferences on Superconducting Quantum Interference Devices and their Applications* (eds Hahlbohm, H. D. & Lübbig, H.) (de Gruyter, Berlin, 1977, 1980, 1985).
- Clarke, J. *Physics Today* **39**, 36-44 (1986).
- Josephson, B. D. *Phys. Lett.* **1**, 251-252 (1962).
- IBM J. Res. Dev.* **24**, 107-264 (1980).
- Ketchen, M. B., Herrell, D. J. & Anderson, C. J. *J. appl. Phys.* **57**, 2550-2574 (1985).
- Rowell, J. M., Gurvitch, M. & Geerk, J. *Phys. Rev.* **B24**, 2278-2281 (1981).
- Anderson, P. W. & Dayem, A. H. *Phys. Rev. Lett.* **13**, 195-197 (1964).
- Ralston, R. W. *IEEE Trans. Magn.* MAG-21, 181-185 (1985).
- Withers, B. S., Anderson, A. C., Green, J. B. & Reible, S. A. *IEEE Trans. Magn.* MAG-21, 186-192 (1985).
- Reible, S. A. *IEEE Trans. Magn.* MAG-21, 193-196 (1985).
- Faris, S. M. *Appl. Phys. Lett.* **36**, 1005-1007 (1980).
- Tuckerman, D. B. *Appl. Phys. Lett.* **36**, 1008-1010 (1980).
- Wolf, P., Van Zeghbroeck, B. J. & Deutsch, U. *IEEE Trans. Magn.* MAG-21 226-229 (1985).
- Gallagher, W. J. *et al. Appl. Phys. Lett.* **50**, 350-352 (1987).
- Hamilton, C. A., Lloyd, F. L. & Kautz, R. L. *IEEE Trans. Magn.* MAG-21, 197-199 (1985).
- Petersen, D. A. *et al. IEEE Trans. Magn.* MAG-21, 200-203 (1985).
- Parker, W. H., Langenberg, D. N., Denenstein, A. & Taylor, B. N. *Phys. Rev.* **177**, 639-664 (1969).
- Kautz, R. L., Hamilton, C. A. & Lloyd, F. L. *IEEE Trans. Magn.* MAG-23, 883-890 (1987).
- Tucker, J. R. & Feldman, M. J. *Rev. mod. Phys.* **57**, 1055-1113 (1985).
- Rudner, S. & Claeson, T. *Proceedings of the International Conference on Superconducting Quantum Interference Devices and their Applications* (eds Hahlbohm, H. D. & Lübbig, H.) 963-985 (de Gruyter, Berlin, 1985).
- Gundlach, K. H., Blundell, R., Ibruegger, J. & Blum, E. J. *Proceedings of the International Conference on Superconducting Quantum Interference Devices and their Applications*, 987-997 (1985).
- Face, D. W., Prober, D. E., McGrath, W. R. & Richards, P. L. *Appl. Phys. Lett.* **48**, 1098-1100 (1986).
- Clarke, J., Hoffer, G. I., Richards, P. L. & Yeh, N.-H. *J. appl. Phys.* **48**, 4865-4879 (1977).
- Clarke, J. *Superconductor Applications: SQUIDs and Machines* (eds Schwartz, B. B. & Foner, S.) 67-124 (Plenum, New York, 1977).
- Clarke, J. *IEEE Trans. Electron. Devices* ED-27, 1896-1908 (1980).
- Ketchen, M. B. & Jaycox, J. M. *Appl. Phys. Lett.* **40**, 736-738 (1982).
- Cromar, M. W. & Carelli, P. *Appl. Phys. Lett.* **38**, 723-725 (1981).
- Voss, R. F. *et al. IEEE Trans. Magn.* MAG-17, 395-399 (1981).
- Van Harlingen, D. J., Koch, R. H. & Clarke, J. *Appl. Phys. Lett.* **41**, 197-199 (1982).
- Tesche, C. D. & Clarke, J. *J. Low Temp. Phys.* **29**, 301-331 (1977).
- Pegrum, C. M. & Donaldson, G. B. *International Superconductivity Electronics Conf.*, Tokyo, 1987.
- Rogers, C. T. & Burhman, R. A. *Phys. Rev. Lett.* **53**, 1272-1275 (1984).
- Tesche, C. D. *et al. Proc. 17th Int. Conf. on Low Temperature Physics* (eds Eckern, U., Schmid, A., Weber, W. & Wuhl, H.) 263-264 (North-Holland, New York, 1984).
- Foglietti, V. *et al. Appl. Phys. Lett.* **49**, 1393-1395 (1986).
- Koch, R. H. *et al. J. Low Temp. Phys.* **51**, 207-224 (1983).
- Colclough, M. S. *et al. Nature* **328**, 47-48 (1987).
- Pegrum, C. M., Donaldson, G. B., Carr, A. H. & Hendry, A. *Appl. Phys. Lett.* **51**, 1364-1366 (1987).
- Zimmerman, J. E., Beall, J. A., Cromar, M. W. & Ono, R. H. *Appl. Phys. Lett.* **51**, 617-618 (1987).
- Iguchi, I., Sugishita, A. & Yanagisawa, M. *Jap. J. appl. Phys.* (in the press).
- Koch, R. H., Umbach, C. P., Clark, G. J., Chaudhari, P. & Laibowitz, R. B. *Appl. Phys. Lett.* **51**, 200-202 (1987).
- Nakane, H., Tarutani, Y., Nishino, T., Yamada, H. & Kawabe, U. *Jap. J. appl. Phys.* **26**, L1925-L1926 (1987).
- Hayakawa, H. *Physics Today* **39**, 46-52 (1986).
- Richards, P. L. *Physics Today* **39**, 54-62 (1986).
- Ferrari, M. J., Johnson, M., Wellstood, F. C., Clarke, J., Rosenthal, P., Hammond, R. & Beasley, M. R., March Meeting of the American Physical Society, New Orleans, LA, March 23, 1988.

The melting history of the late Pleistocene Antarctic ice sheet

M. Nakada & K. Lambeck

Research School of Earth Sciences, GPO Box 4, Australian National University, Canberra 2601, Australia

Spatial and temporal variations in the sea levels of the past 20,000 years around the globe place constraints on the melting history of the major late Pleistocene ice sheets. The Antarctic ice sheets provided a significant contribution to the sea-level rise at a rate that was approximately synchronous with the melting of the Laurentide ice sheet, except for the interval 9,000–6,000 years ago, when it may have lagged behind. Minor melting of the Antarctic ice sheet continued throughout the Holocene.

LATE Pleistocene and Holocene variations in sea level at continental margins and oceanic islands reflect the melting of the ice sheets, the adjustment of the Earth's crust to the redistribution of the ice and water surface loads, and tectonic processes that lead to uplift or subsidence of the crust. Observations of the temporal and spatial patterns of sea-level change therefore contain information on the history of the past ice sheets, on the rheology of the Earth and on tectonic processes, and by a careful examination of sea-levels at different epochs and at different geographical locations it becomes possible to separate some of the parameters that define these contributions.

In many studies (for example, ref. 1) of sea-level change, the principal assumed unknowns have been the parameters defining the Earth's response. Uncertainties in the ice models are, however, sufficiently large to risk erroneous conclusions being drawn about this response^{2,3}, and attempts have been made to constrain both the ice load and the Earth's response from the observed sea level^{1,4,5}. Most of the previous studies have emphasized the importance of the Arctic ice sheets, although Nakiboglu *et al.*⁵ observed that the sea levels in the Pacific region required significant Antarctic melting at about the same time as the Arctic melting after ~20,000 years ago. Here we examine further the question of whether the melting histories of the major ice sheets can be constrained by observations of different parts of the sea-level curve: the levels at the time of maximum glaciation ~18,000–20,000 years ago; the rapid rise of sea-level from ~12,000 to 6,000 years ago and the position of sea level at the end of glacial melting. These three parts of the sea-level curve constrain the late Pleistocene melting history in different ways, depending on whether the observations are from sites near the margins of the former ice sheet (the near-field) or from locations that lie at considerable distance from these loads (the far-field).

Methods

If the evolution of the ice sheets through time and the rheological parameters of the Earth are known, it becomes possible to compute sea-level changes by forward modelling methods^{2,6,7}. The computational scheme and rheological model of Nakada and Lambeck² is used here, in which the Earth is described as a Maxwell solid with a newtonian viscosity that may be depth dependent. Our earlier studies^{2,3} indicated that a very high spatial resolution of the shape of the ocean-continent geometry is required in the neighbourhood of the sites at which sea-level observations are made and an ocean function of 10' resolution has been employed. A consequence of the need for the high spatial resolution is that mid to late Holocene sea-level variations can be quite variable along coastlines over distances of only a few hundred kilometres in special circumstances such as narrow gulfs and bays. This variability has not been noted in previously published global studies of sea-level change. A second consequence is that sea-level change at islands is a function of island size⁸ and only small islands act as 'dipsticks'. Likewise,

Table 1 Estimates of the lowest stand of sea level in the far-field

Area	Ref.	Observed depth (m)	Prediction
			ARC 1 + ANT 1 + BKS (m)
North Queensland	28	114–133	129
Huon Peninsula	9	130	126
Central Japan	29	~130	122
Timor Sea	30	130	136
Central Queensland	31	150–175	130
Arafura Sea	32	150–175	138

The predicted values are based on a mantle with uniform viscosity of 10^{21} Pa s and a lithospheric thickness of 50 km.

a high spatial resolution description of the ice load is required to ensure realistic predictions of sea-level change at sites in the near-field and here we use a 1° representation of the ice load at different epochs. Realistic ice models with such spatial resolution are generally not available and this makes data from sites near the ice margin of limited value for studying the Earth's rheology. Previous studies^{2,5} have shown that at far-field sites the rise in sea level during the late-glacial phase is not particularly sensitive to the choice of rheological model within a geophysically plausible range of values (including a depth-dependent viscosity), and a uniform mantle viscosity of 10^{21} Pa s with a 50-km thick and 10^{25} Pa s viscosity lithosphere is adopted here. Far-field sea levels during the past 6,000 years are more sensitive to rheological parameters because these changes are largely controlled by the Earth's response to the surface loading of the released meltwater.

Observations of sea-level change

Observations of sea-level change earlier than 6,000 yr BP are available from a number of far-field sites although they are not as abundant as the observations of sea-level change for the past 6,000 years. The changes during this earlier period are almost two orders of magnitude greater than those occurring in the latter interval and many of the issues associated with interpreting the evidence for the latter are less critical in establishing the behaviour of sea level during the earlier period. Nevertheless, considerable uncertainty in interpreting the evidence remains.

At only a few sites do the sea-level records extend further back than ~10,000 years and the observations do not constrain the deglaciation history before that time. One exception to this is a constraint on the total volume of meltwater that can be obtained from the observations of the minimum sea-levels at 18,000–20,000 yr BP (Table 1). Generally these levels were 115–150 m below the present level, with 130 m representing a 'reasonable' value^{9,10}. Some regional variation in this level can be expected because of the need to maintain a sea surface that is an equipotential (Table 1).

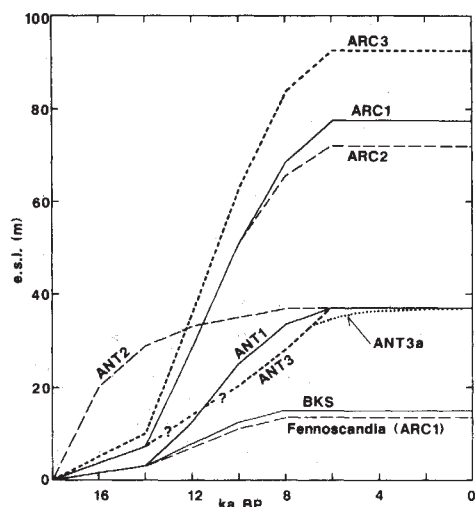


Fig. 1 Equivalent sea-level (e.s.l.) rise as a function of time for the various models for the melting of the ice caps discussed in the text, including the Fennoscandia contribution to ARC 1.

Figure 2 shows observed sea levels at sites near the centre of the former northern ice sheets, and these are believed to be representative of the regions in question for the past 10,000 years. The late glacial far-field sea-levels are from Japan^{11,12}, from the southern Malay Peninsula¹³, from eastern Australia¹⁴ and from New Zealand¹⁵.

Ice models

The starting model for the Arctic ice sheet for the Laurentide and Fennoscandian regions is a 1° linearly interpolated version of the model ICE 1 of Peltier and Andrews⁷. This model is referred to here as ARC 1 (ref. 2) and it defines the ice load at a number of time epochs from 18,000 to 6,000 yr BP. A second model, ARC 2, corresponds to a 1° interpolation of ICE 2 of Wu and Peltier¹ and represents an adjustment of ICE 1 to give better agreement between observed and predicted sea levels near the eastern margin of the Laurentide ice sheet. Differences in predicted sea-level variations between ARC 1 and ARC 2 are significant near the ice margins but not in the far-field. Figure 1 shows the rise in equivalent sea level (e.s.l.), defined as (meltwater volume) × (ice density/seawater density)/(surface area of the oceans) for these two models as a function of time. Table 2 summarizes the total sea-level changes for the two ice models, and they represent only ~55–60% of the observed 130 m. The ARC 1 Laurentide contribution of 59 m compares with 62–78 m estimated by Denton and Hughes¹⁶ for their minimum-ice model reconstruction and with 77–93 m for their

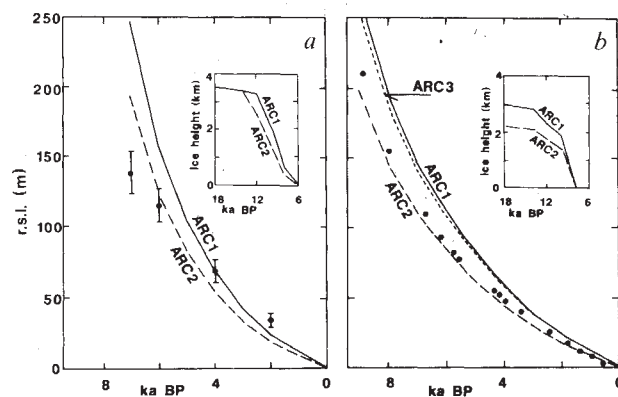


Fig. 2 Observed relative sea-level (r.s.l.) change at two sites in the near-field and the predicted sea-level rise due to the Arctic ice models for a uniform mantle viscosity of 10^{21} Pa s and a 50-km-thick lithosphere. Cape Henrietta Maria (a) lies within the Hudson Bay near the centre of the Laurentide ice sheet³³ and the Angermann River (b) site lies at the northern margin of the Gulf of Bothnia in Fennoscandia. The insets illustrate the melting of the ice sheets at locations near the points of observation.

maximum-ice model, the difference reflecting the generally greater thickness of the central part of the ice sheet for the two models. The Cordilleran, Greenland and Innuitian ice sheets contribute from ~2 to 15 m to the e.s.l. and mountain glaciers contribute up to ~5 m (Table 2).

In so far as the different ice models are based on similar observations, the differences in ice volumes reflect different assumptions made in reconstructing the ice models, on the extent of the bedrock depression beneath the ice load, on the growth and ablation rates of the ice sheet, and on ice mechanics. A particularly important distinction is whether the models are static or dynamic, with the latter models predicting generally greater ice volumes but for shorter time periods¹⁷. The models for the Northern Hemisphere used here are all based on the static-equilibrium assumption. There also remains considerable uncertainty about the timing of the melting of the Laurentide ice sheet. The consensus is that most of the melting occurred from 16,000 to 8,000 yr BP but in some studies it has been suggested that the melting occurred in two phases, the first between 16,000 and 13,000 yr BP and the second between 10,000 and 8,000 yr BP¹⁸. This latter phase follows closely that of the model ARC 1 in Fig. 1, and relative sea levels for the past 8,000 years are little affected by the modification of the melting curve for the earlier period. The contribution of the Fennoscandian ice sheet, including the British Isles, to the total e.s.l. is also quite uncertain but because that ice sheet melted considerably earlier than the Laurentide ice¹⁹, it does not play a major role in determining relative sea-level changes over the past 10,000 years at sites far from the load.

The models ICE 1 and ICE 2 (or ARC 1 and ARC 2) assume that meltwater contributions from the Barents and Kara Seas ice sheets of the eastern Soviet Union were negligible, but Denton and Hughes¹⁶ suggested that a major ice sheet may have existed there in late Pleistocene time. Their model is supported by some glaciologists and geomorphologists²⁰ but others²¹ have suggested that much of the evidence from raised-beach sequences actually corresponds to earlier cycles of deglaciation. Further information is required. Because no observations of sea-level change near this region are used, it suffices to adopt a simple time-dependent disk representation of the Barents-Kara ice sheet whose total volume is equivalent to 15 m of e.s.l. and whose decay is synchronous with the Fennoscandian ice sheet. This, plus ARC 1, comprises the model ARC 3 used below.

A major contribution to e.s.l. is also expected from the Antarctic ice sheet but changes in this ice volume are less well known

Table 2 Sea-level equivalents (metres) of the contributions of the late Pleistocene ice sheets

	ARC 1	ARC 2	ARC 3	From ref. 16 min.	max.
Laurentide	59.0	57.2	59.0	75.6	84.8
Cordillera				0.7	4.6
Greenland (including Innuitian + Iceland)	2.7	2.7	2.7	0.8	9.4
Scandinavia (including British Isles)	15.1	11.5	15.1	19.4	20.1
Barents-Kara	0.7	0.7	15.4	2.1	15.5
Mountain glaciers				4.5	3.2
Total (Northern Hemisphere)	77.5	72.1	92.2	103.1	137.6
Antarctica	ANT 1	ANT 2	ANT 3		
	37.1	37.1	37.1	24.3	24.3

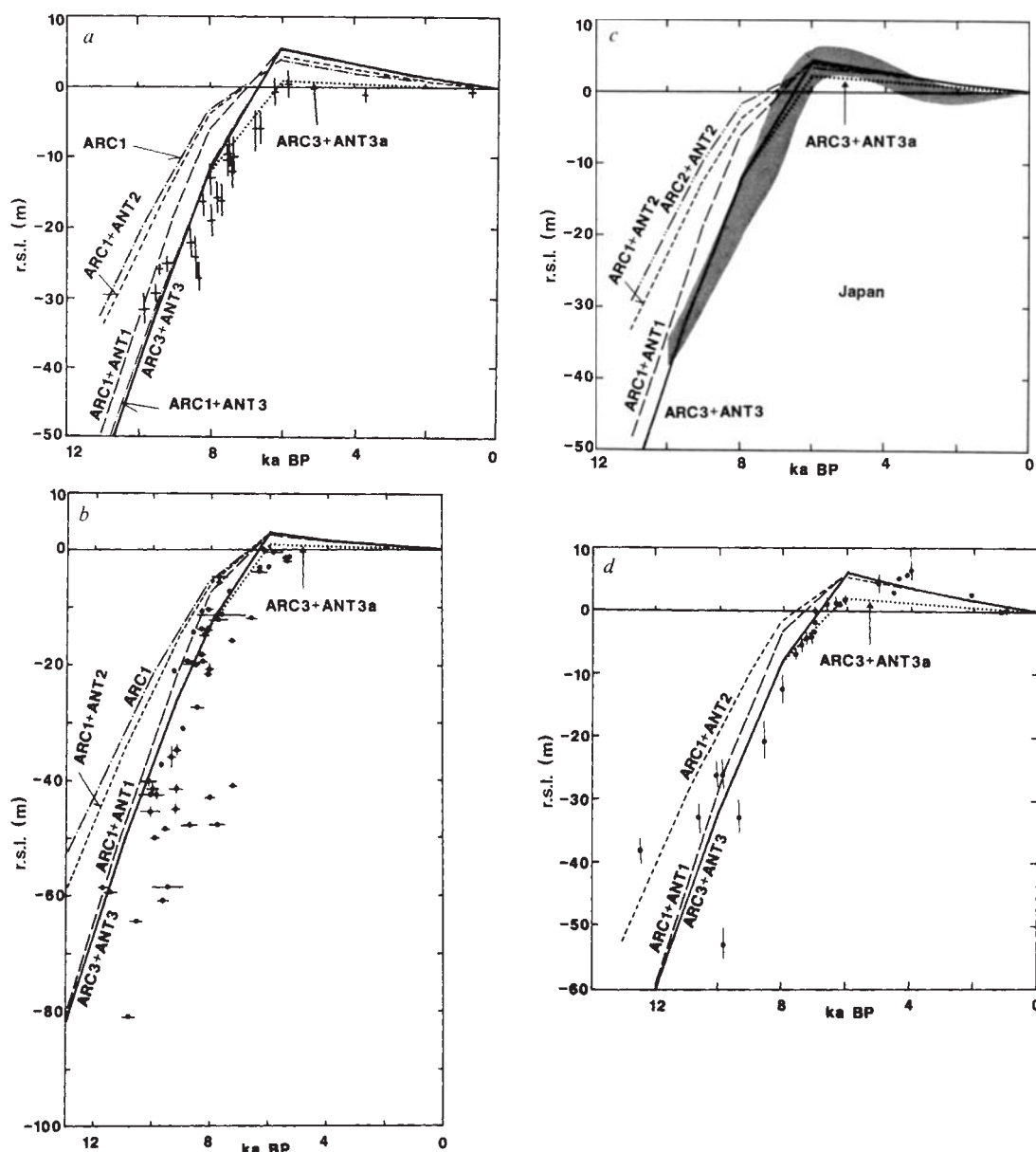


Fig. 3 Observed and predicted sea levels at four far-field sites: *a*, Christchurch, New Zealand; *b*, Moruya, Australia; *c*, Japan; *d*, the Malacca Straits. The Australian and Malacca sites are believed to be tectonically stable^{25,34} whereas the results for the other two sites have been corrected for tectonic movement by using the observed height of the last interglacial sea level at 120,000 yr BP (refs 12, 15).

than their northern counterparts because the ocean-continent geometry has not permitted extensive preservation of either past ice-sheet boundaries or of crustal rebound. As an initial model, ANT 1, we have adopted the difference between the maximum late Pleistocene reconstruction of Denton and Hughes (ref. 16, p. 269) and the present ice thickness given by Drewry²², and this results in an e.s.l. of 37 m. This represents an upper limit because only a part of the grounded ice sheets at the margins of the continent, particularly in the Ross and Weddell Seas, contribute to a change in sea level. Trial calculations have shown that for the sites considered here, all in the Antarctic far-field, the geographic distribution of the Antarctic ice load is not very important for the period before about 6,000 yr BP and we adopt the above model on the basis that it leads to a total sea-level rise of ~130 m, in agreement with the geological evidence (Table 1).

Of greater importance is the timing of the melting of the Antarctic ice sheet, and there are different views as to whether this melting preceded or followed the Arctic deglaciation²³. In ANT 1 the melting is assumed to be synchronous with that of the Arctic ice sheet (see Fig. 1). Wu and Peltier¹ adopted an early melting model (referred to here as ANT 2) in which ~80% of the meltwater was added to the oceans by 14,000 yr BP. Sea

levels for the past 10,000 years at sites away from Antarctica would be little affected by this contribution but with the model ANT 1, 80% of the meltwater was added only by 9,000 yr BP and the Antarctic contributions are reflected in the sea-level change over the past 10,000 years.

Results

Figure 2 illustrates comparisons of predicted and observed sea levels at two sites near the centres of the former northern ice sheets. At both sites the relative sea-level change is dominated by the Earth's response to the unloading of the crust, and the model predictions are quite sensitive to the choice of rheology for the planet. Because the sites lie near the centres of the maximum ice loads these sea levels are not strongly dependent on the details of the edges of the ice sheets.

For a whole range of plausible Earth models, the ARC 1 model overestimates the sea-level change at these and similar sites whereas the ARC 2 model gives better agreement. ARC 2 contains less meltwater than ARC 1 after ~14,000 yr BP and it seems implausible that the overall ice volumes of the eastern and northern Laurentide and the Fenoscandian ice sheets can be significantly increased unless this additional ice melted at an early stage, well before ~12,000 yr BP. The model ARC 3

(which comprises ARC 1 plus the Barents–Kara ice sheet) does not significantly change the sea levels at these sites, and even for Angermann River this latter contribution does not modify the sea levels at these sites for the past 10,000 years by more than ~5%.

Figure 3 shows sea levels at four far-field sites. Here the predicted rises in level up to ~6,000 yr BP are quite insensitive to the choice of rheological model, and these observations place global constraints on the ice models. The models ARC 1 or ARC 2 alone result in a major underestimation of the sea-level change in the interval up to 6,000 yr BP and there is a need to add considerable meltwater from other sources, ranging from ~25 m e.s.l. at 12,000 yr BP to 9 m e.s.l. at 8,000 yr BP. The Antarctic ice model ANT 2 (ref. 1) does not change this discrepancy and if there are strong arguments for an early Antarctic melting history then the volume of Arctic meltwater must be increased or the Arctic melting must be delayed, but this is ruled out by the near-field data of the type given in Fig. 2. Model ANT 1, whose melt history is synchronous with the Arctic melting, leads to a much improved result and the combination of this with ARC 1 gives generally satisfactory agreement with observations after ~12,000 yr BP. The addition of the Barents–Kara ice sheet (that is, ARC 3+ANT 1) does not change this result to any degree because the observational record does not significantly constrain the melting history before about 10,000 yr BP.

Observations at the far-field sites suggest that the predicted rate of sea-level rise from ~9,000 to 6,000 yr BP exceeds the observed rates, and two general adjustments of the ice models can achieve this. One is to retard the melting of the Arctic ice sheet at this time by ~1,000 years but this is inconsistent with the sea levels near the centre of the Arctic ice sheet. The alternative is to bring the melting of the Antarctic ice sheet forward by ~1,000 years, and a possible mean melting model is illustrated in Fig. 1 as ANT 3. No constraints on this model can be placed for times before ~10,000 yr BP.

All ice models so far discussed predict far-field sea-level maxima at 6,000 yr BP that are considerably higher than observed. The typical observations are of levels ~0–2 m above the present level, whereas the predictions are up to 6 m above this for a wide range of mantle viscosity models (e.g. Figs 3 and 4). A second discrepancy occurs at island sites where the observed Holocene maximum levels often occur at ~4,000 yr BP²⁴ rather than at the predicted 6,000 yr BP. Part of the explanation for these discrepancies could be that the sites illustrated in Figs 3 and 4 are subject to tectonic subsidence, but there is either no independent geological evidence for this or, in the case of the Japan and Christchurch data, corrections for this have been made. The island observation could be attributed to a time lag in the reef's response to sea-level change, but different sea-level indicators produce similar results. The introduction of Earth models with very thick elastic lithospheres does partly, but not consistently, reduce these high levels, although it does not resolve the question of the time lag in the island sea-level response. Such models can be tested by examining relative sea levels in bays and gulfs where the sea levels are sensitive to the geometry of the distribution of the meltwater in the neighbourhood of the sites, for here the thick lithosphere models dampen out any regional variation in the response much more than do thin lithosphere models. Evidence of this kind is observed around Australia's coastlines including the Gulf of Carpentaria and Spencer Gulf. The conclusion reached is that the differential sea levels are consistent with models of a lithospheric thickness of ~50 km and that the observed values of the absolute amplitudes of the Holocene sea levels are indicative of continuing addition of meltwater into the oceans during the Holocene amounting to ~3 m e.s.l. (M.N. and K.L., in preparation). The obvious source is from Antarctica because there is no reason why melting of the Antarctic ice sheet ceased abruptly at 6,000 yr BP, in particular if the melting of this ice sheet lagged behind the melting of the northern ice sheets during the early

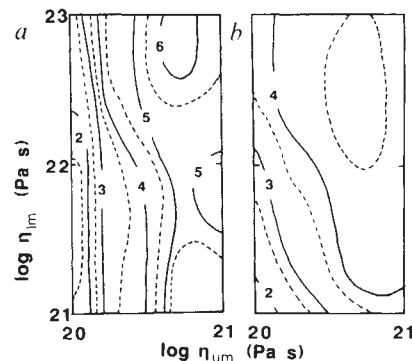


Fig. 4 Sea-level (in metres) at 6,000 yr BP above present level as a function of upper (η_{um}) and lower (η_{lm}) mantle viscosity at *a*, Christchurch and *b*, Moruya. The viscosity discontinuity is taken to be at 670 km depth. The ice load is ARC 3+ANT 3.

Holocene. This is the model ANT 3a in Fig. 1. An alternate source would be mountain glaciers, and the sea-level data alone, in the absence of data near the edge of the Antarctic ice sheet, cannot discriminate between these two alternate hypothesis. But it is improbable that the mountain glaciers, whose maximum Pleistocene ice volume is equivalent to only ~5 m (ref. 16), could have stored this additional ice. Figure 3 shows the predicted far-field sea-level change for the preferred ice model ARC 3+ANT 3a for a two-layered mantle viscosity (2×10^{20} Pa s for the upper mantle and 10^{22} Pa s for the lower mantle). This particular combination of ice and mantle models satisfies the majority of the far-field sea-level observations from tectonically stable sites.

Discussion

The examples discussed above illustrate the potential of past sea-level observations to constrain models of the melting of the ice sheets in late Pleistocene time, although the currently available observations are inadequate for determining unique solutions of the melt histories. The conclusions, in order of decreasing certainty, are as follows:

- (1) In addition to the Northern Hemispheric ice sheets of Laurentia and Fennoscandia, as modelled by Denton and Hughes¹⁶, there must be other sources of meltwater equivalent to ~55 m of e.s.l.
- (2) The missing meltwater source cannot lie in the eastern and northern part of the Laurentide or the Fennoscandian ice sheets unless melting occurred considerably earlier than about 10,000 yr BP. Thus if, as suggested by the dynamic models the Laurentide ice volume was greater than that used here, the increased volume could not have existed after ~12,000–10,000 yr BP.
- (3) The relative sea-level observations of the past 10,000 years do not constrain the assumption that the Barents–Kara ice sheets made a significant contribution to the sea-level rise at about the same time as the melting of the Fennoscandian ice sheet, but such a contribution does help in explaining the sea levels of ~130 m below present level during the time of the last glacial maximum.
- (4) The melting of the Antarctic ice sheet provided a significant contribution to the late Pleistocene sea-level rise, perhaps 35 m e.s.l. at a rate that was synchronous with, or lagged behind, the decay of the Arctic ice sheet, particularly between ~9,000 and 6,000 yr BP.
- (5) An amount of meltwater, equivalent to ~3 m e.s.l., was added to the oceans over the past 6,000 years.

These conclusions are interdependent. Major changes in the northern ice sheet models, for example, have immediate repercussions on conclusions (3) and (4), and additional observations of sea level are required, particularly from sites near the previously glaciated regions. Unambiguous observations of relative

sea-level change in the neighbourhood of Franz Josef Land and Nova Zemlya, for example, would help constrain the Barentz and Kara ice sheets. Observations along the north-west coast of Canada or along the northern margin of Alaska would contribute to understanding the contributions of the western parts of the Laurentide and the Cordilleran ice sheets to the global sea-level change. Observations closer to Antarctica, particularly quantitative uplift estimates for some of the elevated terraces found on that continent, would contribute significantly. Additional observations for the early part of the sea-level rise in late Pleistocene time in the far-field would help to constrain both the total volume of meltwater and the rate at which this was added to the oceans.

The conclusions drawn above have consequences of glaciological significance. One is that the late Pleistocene ice sheets may

have held more ice than currently thought. It does not appear that the ice volumes of the Laurentide and Fennoscandian ice sheets can be significantly increased after ~12,000–10,000 yr BP and that these ice sheets can contribute only ~75 m of the 130 m required by the sea levels during the last glacial maximum^{10,25}. Some of this, perhaps 15 m, may have come from the Barents–Kara ice sheets, but Antarctica is likely to have contributed a significant fraction of the total, perhaps 50% greater than estimates by Denton and Hughes. The conclusion concerning the timing of the Antarctic meltwater lends support to the hypothesis that the Antarctic ice sheets melted in response to the rising sea level caused by the decay of the northern ice sheets^{26,27}, and this is further supported by the suggestion that the addition of meltwater to the oceans continued after melting ceased in the Northern Hemisphere.

Received 10 November 1987; accepted 7 March 1988.

1. Wu, P. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **74**, 377–449 (1983).
2. Nakada, M. & Lambeck, K. *Geophys. J. R. astr. Soc.* **90**, 171–224 (1987).
3. Nakada, M. & Lambeck, K. in *Mathematical Geophysics* (eds Vlaar, N. J., Nolet, G., Wortel, M. J. R. & Cloetingh, S. A. P. L.) 347–361 (Reidel, Dordrecht, 1988).
4. Quinlan, G. & Beaumont, C. *Can. J. Earth Sci.* **18**, 1146–1163 (1981).
5. Nakiboglu, S. M., Lambeck, K. & Aharon, P. *Tectonophysics* **91**, 335–358 (1983).
6. Farrell, W. E. & Clark, J. A. *Geophys. J. R. astr. Soc.* **46**, 647–667 (1976).
7. Peltier, W. R. & Andrews, J. T. *Geophys. J. R. astr. Soc.* **46**, 605–646 (1976).
8. Nakada, M. *Tectonophysics* **121**, 263–276 (1986).
9. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
10. Chappell, J. *Quat. Res.* **4**, 405–428 (1974).
11. Sugimura, A. *Kagaku* **47**, 749–755 (1977).
12. Yonekura, N. & Ota, Y. *Rec. Prog. nat. Sci. Japan* **11**, 17–34 (1986).
13. Geyh, M. A., Kudrass, H. R. & Streif, H. *Nature* **278**, 441–443 (1979).
14. Thom, B. G. & Roy, P. S. in *Australian Sea Levels in Last 15,000 Years: A Review* (ed. Hopley, D.) 64–84 (James Cook University of North Queensland, 1983).
15. Gibb, J. G. *Bull. R. Soc. N.Z.* **24**, 377–395 (1986).
16. Denton, G. H. & Hughes, T. J. (eds) *The Last Great Ice Sheets* (Wiley, New York, 1981).
17. Budd, W. F. & Smith, I. N. *I.A.H.S. Publ.* **131**, 369–409 (1981).
18. Duplessy, J. C., Delibrias, G., Turon, J. L., Pujol, C. & Duprat, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 121–144 (1981).
19. DeGeer, E. H. *Geol. För. Stockh. Förh.* **76**, 299–329 (1954).
20. Grosswald, M. G. *Quat. Res.* **13**, 1–32 (1980).
21. Boulton, G. S. *Boreas* **8**, 31–57 (1979).
22. Drewry, D. J. (ed.) *Antarctica: Glaciological and Geophysical Folio* (Scott Polar Res. Inst., Cambridge, 1982).
23. Broecker, W. S. in *Milankovitch and Climate Pt 2* (ed. Berger, A. L.) 687–698 (Reidel, Dordrecht, 1984).
24. Pirazzoli, P. A., Montaggioni, L. F., Delibrias, G., Faure, G. & Salvat, B. *Proc. Fifth Int. Coral Reef Congr. Tahiti* **3**, 131–134 (1985).
25. Chappell, J. in *Sea-level Changes* (eds Tooley, M. J. & Shennan, I.) 296–331 (Blackwell, Oxford, 1987).
26. Stuiver, M., Denton, G. H., Hughes, T. J. & Fastook, J. L. in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T. J.) 319–439 (Wiley, New York, 1981).
27. Budd, W. F. & Smith, I. N. *Ann. Glaciol.* **3**, 42–49 (1982).
28. Carter, R. M. & Johnson, D. P. *Mar. Geol.* **71**, 137–164 (1986).
29. Ota, Y., Matsushima, Y. & Moriaki, H. *Atlas of Holocene Sea Level Records in Japan* (Japanese working group of Holocene sea-level project, IGCP, 1981).
30. van Andel, T. M. & Veevers, J. J. *Bur. Min. Res. Geol. Geophys. Aust. Bull.* **83**, 1–173 (1967).
31. Veeh, H. H. & Veevers, J. J. *Nature* **226**, 536–537 (1970).
32. Jongsma, D. *Nature* **228**, 150–151 (1970).
33. Walcott, R. L. A. *Rev. Earth planet. Sci.* **1**, 15–37 (1973).
34. Batchelor, B. C. *Bull. geol. Soc. Malaysia* **11**, 283–313 (1979).

The SV40 enhancer contains two distinct levels of organization

Brian Ondek, Lisa Gloss & Winship Herr

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

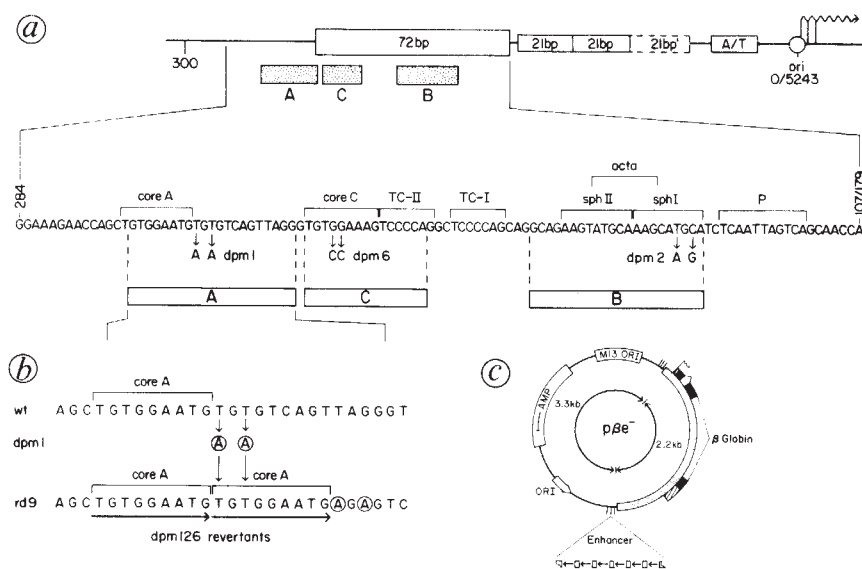
The SV40 transcriptional enhancer is composed of separate 15 to 20 base-pair-(bp)-long enhancer elements that cooperate with one another or duplicates of themselves to enhance transcription. These elements are bipartite, being composed of subunits, called enhansons, that can be duplicated or interchanged to create new enhancer elements. Enhansons differ from the enhancer elements because they are very sensitive to changes in spacing. This prototypic enhancer, therefore, contains two distinct levels of organization, each of which requires redundancy to be effective.

ENHANCERS are *cis*-acting regulatory elements that function independently of distance and position relative to the transcriptional start site. Since their discovery within the regulatory region of the small simian DNA tumour virus SV40 (refs 1, 2), enhancers have been characterized by the presence of multiple short (8–10 bp) sequence motifs. Frequently, these motifs were first identified by nucleotide sequence comparisons between different enhancers and later established as functionally significant by mutational analysis. For example, the 'core' motif GTGG^A/T^A/T^A/T^G (ref. 3) was first identified by sequence comparisons between the SV40 and murine sarcoma virus (MSV) enhancers⁴. Subsequently a systematic mutational analysis of the SV40 enhancer region showed that the core as well as other motifs, such as the sph motif AAG^T/C/ATGCA, are functionally important⁵. Many of the SV40 enhancer sequence motifs appear to represent binding sites for *trans*-acting factors as shown by

using partially purified^{6–13} or fully purified proteins^{14–19}; some of these fully purified factors have been shown to activate transcription *in vitro* although dependent on an upstream position^{15–18}. It thus appears that enhancers are composed of a complex array of sequence motifs that interact with distinct factors.

In a separate experimental approach involving viral genetic selections, the SV40 enhancer has been shown to contain at least three discrete 15–20 bp-long elements, called A, B and C, that cooperate with one another or duplicates of themselves to enhance transcription^{20–22}. These enhancer elements function autonomously when present in two or more copies and display unique patterns of cell-specific enhancer activity^{23,24}. They differ, however, from sequence motifs because mutant elements, containing wild-type motifs known to be important for enhancer function⁵, are inactive²³. Here we show that the A and B enhancer

Fig. 1 SV40 early promoter showing the location of enhancer elements and sequence motifs. *a*, Diagram of the SV40 early promoter showing, from right to left, the early transcriptional start sites, origin of replication. AT-rich TATA-like element, one imperfect (dashed box) and two perfect 21-bp repeats, and a single copy of the wild-type 72-bp repeat. The A, B and C enhancer elements (stippled boxes) are shown below their corresponding position within the early promoter. The nucleotide sequence of the enhancer region (nucleotides 179–284) is shown below with sequence motifs bracketed above the sequence, and the *dpm1*, *dpm2* and *dpm6* point mutations identified below the sequence along with the location of the A, B and C elements. The coreA and coreC motifs TGTGGAA^T/A^G overlap the related CACA motifs³⁰ and are subsets of the GT motifs I and II (G^C/G^TTGTGGAA^T/A^GT) (ref. 5). The TC-I and TC-II, sphI and sphII, and P motifs are described by Zenke *et al.*⁵. The junction of the two sph motifs creates a sequence similar to the octamer (octa) motif found in immunoglobulin gene promoters²⁵. *b*, The sequence of the wild-type A element and flanking region is shown with the coreA motif bracketed. The *dpm1* mutations are circled, and the structure of the *rd9* coreA duplication found in the three *dpm126* revertants, *rd51*, *rd110* and *rd136* (ref. 20), is shown at the bottom. *c*, The human β -globin expression plasmid *pBe⁻* (ref. 23) contains the human β -globin gene (stippled box) within the polylinker (hatch marks) of pUC119. Synthetic enhancers were cloned into the sphI recognition site (labelled 'enhancer') 2.2-Kb downstream or 3.3-Kb upstream of the β -globin transcriptional start site. Synthetic DNA segments (arrows) are shown repeated six times and separated by *Xho*I recognition sites (boxes) in the '+' orientation such that the sequence shown in *a* reads clockwise.



elements are bipartite, containing subunits that correlate with the core and sph sequence motifs. These subunits can be duplicated or combined with a heterologous subunit to create new enhancer elements. But, unlike enhancer elements themselves, these subunits are very sensitive to changes in spacing. These results lead to a picture of the SV40 enhancer as a composite of many subunits, correlating with sequence motifs and protein binding sites, that are organized in a hierarchical structure: first creating enhancer elements and subsequently an enhancer. We refer to these subunits that form enhancer elements as enhansons because they appear to be the basic units of enhancer structure.

Tandem sequence motifs as enhancer elements

Figure 1*a* shows the nucleotide sequence of the SV40 enhancer along with the position of the A, B and C enhancer elements^{20,21}, the mutations (*dpm1*, *dpm2* and *dpm6*) which inactivate these three elements^{20,23}, and sequence motifs shown to be important for SV40 enhancer function⁵. Figure 1*b* shows the structure of a small 9-bp duplication of the coreA motif which led us to test whether duplicated motifs might create functional enhancer elements. The coreA duplication arose in three revertants of the defective SV40 enhancer triple A, B and C element mutant *dpm126* (ref. 20). These revertants each arose by two successive duplications; the first duplication created the 9-bp coreA motif duplication, whereas a second, variably sized duplication, created two copies of the new 18-bp segment. The double duplication pattern suggested that the coreA motif duplication was creating a new enhancer element.

To test this hypothesis, a multimerized synthetic enhancer containing six tandem copies of the coreA motif duplication (coreA/coreA), separated by *Xho*I recognition sites, was constructed as described in Fig. 2 and cloned 2.2-Kb downstream of the human β -globin transcriptional start site in the expression vector *pBe⁻* (ref. 23) shown in Fig. 1*c*. Figure 2*a* shows the result of an assay of β -globin RNA expression with the coreA/coreA construct after transfection of CV-1 cells, the African Green Monkey kidney cell line used for the SV40 revertant isolations. In Fig. 3, the structure of synthetic enhancer elements is shown along with the relative activity of each corresponding multimerized enhancer in CV-1 cells. As described

previously²³, multiple isolated coreA motifs are inactive (A17, Fig. 2*a*, lane 2), but in the context of the full A element (A21, lane 3) the coreA motif is active. Unlike the isolated coreA motifs of A17, the tandem coreA motifs in coreA/coreA are exceptionally active as an enhancer (lane 4). To demonstrate that both 9-bp coreA motifs are required for coreA/coreA enhancer activity, closely spaced point mutations analogous to those of *dpm6* in the coreC motif (Fig. 1*a*) were introduced into each sequence and shown to inactivate the coreA/coreA element (Fig. 2*a*, lanes 5 and 6). Thus, tandem but not isolated coreA motifs can enhance transcription.

To test the activities of other repeated SV40 enhancer sequence motifs, we analysed the significance of the single base difference between the two 9-bp sph motifs which form the majority of the B enhancer element (see Figs 1*a* and 3*c*). Together the two sph motifs are active as an enhancer when present in multiple copies but the *dpm2* mutations within the sphI motif inactivate the entire element (ref. 23 and Fig. 3*c*). Figure 2*b* shows that, when assayed in a multimerized enhancer construct, changing the wild-type sphII/sphI sequence (lane 2) to sphII/sphII (lane 3) generates a slightly better enhancer, but changing the sequence to sphI/sphI (lane 4) has a large deleterious effect on enhancer function. Nevertheless the ineffective sphI/sphI construct can be reactivated by changing the sequence to an sphI/sphII configuration (lane 5) in which the sph motifs are inverted compared to the wild-type sphII/sphI construct. This experiment shows that the two sph motifs are not equivalent and suggests the sphI motif is a weak subunit. These results are consistent with the hypothesis that the B element sphI and sphII motifs are discrete subunits that cooperate with one another to enhance transcription.

Like the sph motifs, the SV40 enhancer coreA and coreC motifs differ at a single position; the coreC motif contains a single T to A transversion at position 8 of the coreA motif, which is permitted by the consensus core sequence GTGG^A/T^A/T^A/T^G (ref. 3). Figure 2*c* shows, however, that combinations of the coreA and coreC motifs (lanes 3 and 4) or two coreC motifs together (lane 5) are inactive as enhancers in CV-1 cells. These results show that the highly active coreA/coreA construct can be very sensitive to single point

Fig. 2 Sequence motif duplications and combinations create new enhancer elements. The levels of α - and β -globin probe RNAs protected from RNase digestion, by cytoplasmic RNA isolated from transfected African Green Monkey kidney CV-1 cells, are shown after separation by electrophoresis through denaturing 6% polyacrylamide gels. The correctly initiated α -globin (132 nucleotides), which serves as a reference for transfection efficiency, and the experimental β -globin (350 nucleotides) protected fragments are labelled α and β , respectively. The p β plasmids with (1X72) and without (enh^-) one copy of the 72-bp repeat²³ are shown for comparison. Each lane represents an assay of a p β construction containing six tandem copies of the oligonucleotide indicated above the lane. The sequence of each repeat and quantitation of enhancer activity are shown in Fig. 3. *a*, A element and coreA motif constructs; *b*, sph motif constructs; and *c*, coreA and coreC motif combinations and sphII/coreA chimera.

Methods. The synthetic enhancers were constructed from complementary synthetic oligonucleotides as described previously²³. The sequence of each pair of oligonucleotides was $\text{GN}_{18}\text{CTCGA}$, where N_{18} is the 18 nucleotide-long unique sequence shown in Fig. 3, and its complement $\text{AGN}_{18}\text{CTCG}$, which when annealed to one another produced a 3-base 3' overhang. The intervening *Xho*I recognition sequence CTCGAG was included to buffer each test element with a constant flanking sequence. Briefly, the synthetic enhancers were constructed by ligation of annealed oligonucleotides, followed by end-repair and polyacrylamide gel electrophoretic purification of fragments containing six tandem copies. The purified fragments were blunt-end ligated into the *Sph*I site of p βe^- , which had been end-repaired by treatment with the large fragment of DNA polymerase I. Potential clones were screened by DNA sequencing. Plasmid DNAs were transfected into CV-1 cells by calcium phosphate co-precipitation; cytoplasmic RNAs were collected 48 hours post-transfection and analysed by RNase protection of internally-labelled single-stranded α - and β -globin RNA probes as described previously²³.

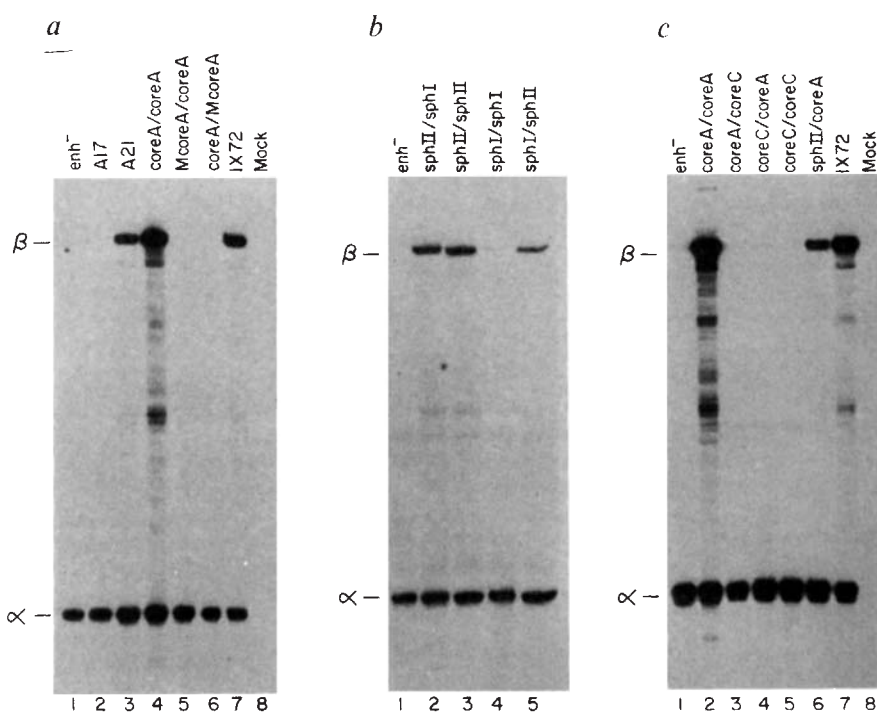


Fig. 3 Structure of synthetic enhancer repeats and quantitation of enhancer activity. The nucleotide sequence of each multimerized repeat is shown with the levels of β -globin expression in CV-1 cells relative to the wild-type 1X72 construct shown to the right. Dashed lines and arrows point to the changes each repeat contains with respect to their parent repeat shown above. For example, the sphI/sphI repeat has a single T $\rightarrow$ C transition with respect to the sphII/sphI parent repeat. The *Xho*I recognition sequences (CTCGAG) which join each repeat are shown boxed. The sequence of the oligonucleotides and the structure of each multimerized enhancer construct containing six copies of each repeat is described in Fig. 2 and Fig. 1 legends, respectively. *a*, The wild-type enhancer constructs containing one and two copies of the 72-bp repeat (1X72, 2X72), the enhancerless control (enh^-), and the wild-type 17- and 21-bp A element constructions (6XA17 and 6XA21) were as described²³. The A17 and A21 constructions do not contain *Xho*I restriction sites and when ligated in tandem match the wild-type sequence by an additional 1 and 3 bp, respectively, as shown by the lower case letters. *b*, The coreA, coreC and chimaeric sphII/coreA containing constructions; *c*, sph motif constructions; and *d*, the coreA/coreA 1-10-bp spacing constructions with the intervening nucleotide sequence shown below the insertion site. The sphII/MspI construction contains the *dpm2* point mutations (see Fig. 1) described previously²¹. The quantitation of enhancer activity was determined by scintillation counting of the correctly initiated α - and β -globin RNAs except for McoreA/coreA and coreA/McoreA which were quantitated by densitometry. A minimum of two and on average five transfections were performed with each construct and a minimum of two transfections were quantitated for each construct. Low levels of enhancerless activity (-) were always less than 3% and generally $\sim$ 1% the activity of 1X72. The synthetic enhancers were in the [+] orientation shown in Fig. 1c except the A17 and A21 constructs in *a*, McoreA/coreA, coreA/McoreA and sphII/coreA constructs in *b*, sphII/MspI construct in *c*, and all constructs (including the coreA/coreA control but not coreA-10-coreA) in *d*, which are positioned in the [-] orientation. In these and previous experiments²³, less than twofold differences in the enhancer activity have been observed between matched constructs with the enhancers positioned in opposite orientations. The coreA/coreA orientation was obtained by inverting the coreA/coreA⁺ insert as described previously²³.

Construct	Oligonucleotide Sequence	Activity (% 1X72)
<i>a</i> 1X72		100
2X72		240
enh^-		(-)
A17	TGTGGAATGTGTGTCAG†	(-)
A21	GTGGAATGTGTGTCAGT TAGG g1g	25
<i>b</i> coreA/coreA	GAG core A core A TGTGGAATGTGTGGAATG CTC	190
McoreA/coreA	GAG core A TGTGGAATGTGTGGAATG ↓ CC CTC	(-)
coreA/McoreA	GAG core A TGTGGAATGTGTGGAATG ↓ CC CTC	(-)
coreA/coreC	GAG core A TGTGGAATGTGTGGAATG ↓ A CTC	(-)
coreC/coreA	GAG core C TGTGGAATGTGTGGAATG ↓ A CTC	(-)
coreC/coreC	GAG core C TGTGGAATGTGTGGAATG ↓ A CTC	(-)
sphII/coreA	GAG sph II AAGTATGCA ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ CTC	35
<i>c</i> sphII/sphI	GAG sph II sph I AAGTATGCA AAGCATGCA CTC	110
sphII/sphII	GAG sph II AAGTATGCA AAGCATGCA ↓ T CTC	150
sphI/sphI	GAG sph I AAGTATGCA AAGCATGCA ↓ C CTC	6
sphI/sphII	GAG sph I AAGTATGCA AAGCATGCA ↓ C CTC	50
sphII/MspI	GAG sph II AAGTATGCA AAGCATGCA ↓ T ↓ A CTC	(-)
<i>d</i> coreA/coreA	GAG core A core A TGTGGAATGTGTGGAATG CTC	120
coreA-1-coreA	GAG core A TGTGGAATGTGTGGAATG ↓ C CTC	7
coreA-2-coreA	GAG core A TGTGGAATGTGTGGAATG ↓ AC CTC	7
coreA-5-coreA	GAG core A TGTGGAATGTGTGGAATG ↓ ACATA CTC	(-)
coreA-10-coreA	GAG core A TGTGGAATGTGTGGAATG ↓ GATGATCGAG CTC	(-)

mutations, and that although the coreA and coreC motif sequences are both involved in SV40 enhancer activity⁵, they are functionally distinct. The difference in activity between these two motifs is consistent with their different interactions with nuclear *trans*-acting factors^{10-12,17}.

In Fig. 2, the coreA and sph motif duplications have been assayed in CV-1 cells, the same cell line used to isolate the SV40 enhancer revertants. Figure 4 shows that the elements that are active in CV-1 cells are also functional in HeLa cells, a human cervical carcinoma cell line frequently used to isolate SV40 enhancer binding factors. As described previously for the A, B and C elements^{23,24}, there may exist differences in the cell-specific activity of these new enhancer elements because the coreA/coreA (lane 2) enhancer and the sphII/coreA (lane 10) enhancer described below are somewhat less active in HeLa cells relative to the sph motif constructs.

Sub-elements are interchangeable

The coreA and sph motifs behave as discrete subunits that function in pairs to form active enhancer elements. If the A and B enhancer elements are indeed bipartite it may be possible to mix subunits between elements to create new chimeric elements. To test this hypothesis, a B/A element chimaera was designed such that the A element coreA motif is juxtaposed with the heterologous B element sphII motif. In this sphII/coreA chimaera every position within the left coreA motif of the coreA/coreA construct is mutated (see Fig. 3b). Nevertheless, this chimaeric B/A element, when multimerized, is active as an enhancer (Fig. 2c; lane 6). The activity exhibited by this chimaeric enhancer element (35% of 1X72) contrasts dramatically with the lack of activity of the single and double point mutants of the coreA and sph motif constructions. Therefore, motifs taken from separate wild-type enhancer elements can be combined to form a new functional enhancer element. Taken together the results of the various motif constructions shown in Fig. 3 are consistent with a model in which SV40 enhancer elements are structurally bipartite, consisting of subunits that correlate both with motifs found in other enhancers and with binding sites for *trans*-acting factors. We suggest that these subunits be called 'enhansons', representing the basic unit of enhancer structure.

Enhanson/element spacing constraints

Although tandem coreA 'enhansons' combine to produce a functional enhancer element, multiple copies of well-separated coreA enhansons are not active because the multimerized A17 and mutant coreA/coreA constructs, which contain wild-type coreA enhansons separated by 8 and 15 bp, respectively, are inactive. Similarly, isolated sphII enhansons are also inactive (see sphII/MspI in Fig. 3c). To test directly the effects of spacing between the 9-bp coreA repeats in the coreA/coreA synthetic enhancer constructs, a series of parallel constructions was made separating the tandem coreA motifs by 1, 2, 5 and 10 bp within each repeat as shown in Fig. 3d. The results of one assay of these constructs are shown in Figs. 5a. The 1- and 2-bp insertions (lanes 3 and 4) exhibit enhancer activity as compared to the enhancerless control (lane 1) but considerably less activity than the coreA/coreA construct (lane 2). But the 5- and 10-bp insertions (lanes 5 and 6) are no more active than the enhancerless control. The sharp but not complete loss of activity with the 1- and 2-bp insertions compared to the numerous inactive coreA mutants shown in Figs 2 and 3 suggests that spacing between coreA motifs is critical but not inflexible and that the coreA/coreA junction sequence itself is not entirely responsible for coreA/coreA enhancer activity.

In contrast to the strict enhanson spacing requirements, individual enhancer elements display a relaxed spacing requirement, as shown by the experiment in Fig. 5b in which the effect of spacing introduced between two intact B elements was assayed. There is a gradual decline in activity as spacing is

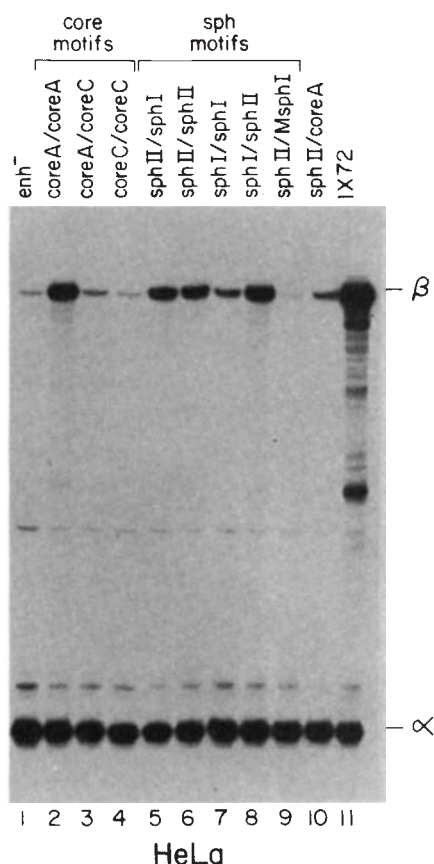


Fig. 4 Activity of the multimerized core and sph motif enhancers in HeLa cells. The result of HeLa cell transfection and assay of α - and β -globin expression is shown with the plasmid construct used for transfection indicated above each lane. See Fig. 3 for the structure of each enhancer. DNAs were transfected by calcium phosphate coprecipitation and RNAs analysed by RNase protection as described in Fig. 2 legend.

increased until at a separation of 125 bp no activity is detected. This emphasizes that individual enhancer elements are ineffective as enhancers. These spacing constructs along with two other insertions of 19 and 20 bp (data not shown) also indicate that helical periodicity plays little, if any, role in enhancer element interaction. The gradual decline in activity with increased spacer DNA between elements compared with the rapid loss of activity with separation of enhansons differentiates elements from enhansons. It also indicates that there are fundamental differences in the way enhansons and enhancer elements cooperate to activate transcription at a distance.

Enhancer substructure

The experiments described here suggest that the SV40 enhancer contains two levels of organization, enhancer elements and sub-elements called enhansons, the latter of which correlate with sequence motifs known to be important for enhancer function⁵. This picture has developed from structural characterization of SV40 revertant enhancers arising when the enhancer is inactivated by point mutation. The initial experiments, in which large tandem duplications restored activity, indicated that the enhancer is composed of multiple elements that when duplicated can compensate for loss of function in heterologous elements²⁰⁻²². The presence of a second organizational level has been revealed by the analysis of complex rearrangements that arose in revertants of the SV40 enhancer triple A, B and C element mutant *dpm126*. Many of these revertants contained short duplications such as the 9-bp coreA motif duplication studied here²⁰. The significance of these short rearrangements was determined by

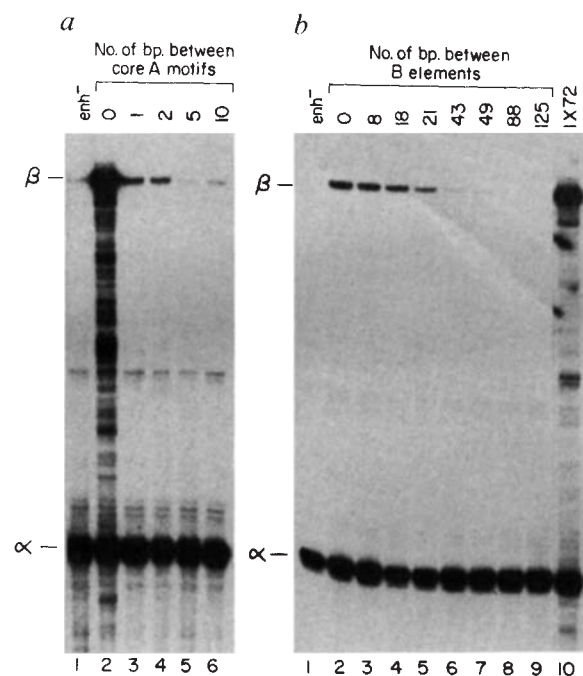


Fig. 5 Active pairs of enhancers and elements respond differently to increased spacing. *a*, The coreA/coreA constructs containing 1, 2, 5 and 10-bp (lanes 3–6) spacing supplements positioned between the coreA motifs (see Fig. 3) were assayed by transfection into CV-1 cells alongside the parental coreA/coreA construct (lane 2). *b*, A plasmid containing two tandem 22-bp B elements (lane 2) and a series of plasmids containing insertions of DNA separating the two B elements by 8–125-bp (lanes 3–9) were tested by transfection into CV-1 cells. In both experiments, levels of α - and β -globin expression were assayed as described in Fig. 2 legend.

Methods. Preparation of B element spacing constructs. The B-element spacing constructions were made starting with the IXB22 construct²³, which underwent separate digestions with either *ScaI* and *PstI*, or *ScaI* and *HindIII*, followed by end-repair with the large fragment of DNA polymerase I and isolation of the two different B-element-containing fragments using low melting agarose. These two fragments were ligated together with either (1) no insert, resulting in an 8-bp insertion (CAAGCTGG) or (2) a *XhoI* linker (underlined) resulting in an 18-bp (CAAGCTCCCTCGAGGGGG) insertion. To prepare the larger spacing constructs the *XhoI* containing construct was digested with *XhoI*, end-repaired, and ligated to random blunt-ended λ phage *Sau3A* restriction fragments. Clones were screened by dideoxynucleotide sequencing using a single chain terminator (T tracking) and clones containing a total of 43, 49, 88 and 125 bp separating the two B elements were isolated and fully sequenced. The 21-bp insertion is the result of self-ligation of the filled-in *XhoI* digested 18-bp construct with the loss of one base pair.

using enhancers created by multimerization of synthetic oligonucleotides so that the activity of an individual element could be independently assayed.

To establish that the activity of the synthetic coreA and sph motif constructs is due to the juxtaposition of two separate sub-elements and not the fortuitous creation of a new site at the junction between motifs, we made mutations within the individual motifs and examined the effects of separating the motifs. Consistent with a bipartite element structure, a point mutation(s), within one of two paired coreA or sphII motifs has the same effect when the mutation is located in either motif (see Fig. 3*b*, *c*). The effect of increased spacing between the coreA motifs is also consistent with a bipartite structure because, unlike the inactive coreA/coreA point mutants (Fig. 3*b*), the 1- and 2-bp insertions are still active (Fig. 3*d*). In designing the 1- and 2-bp insertions we specifically chose sequences that seemed least likely to substitute functional sequences at the

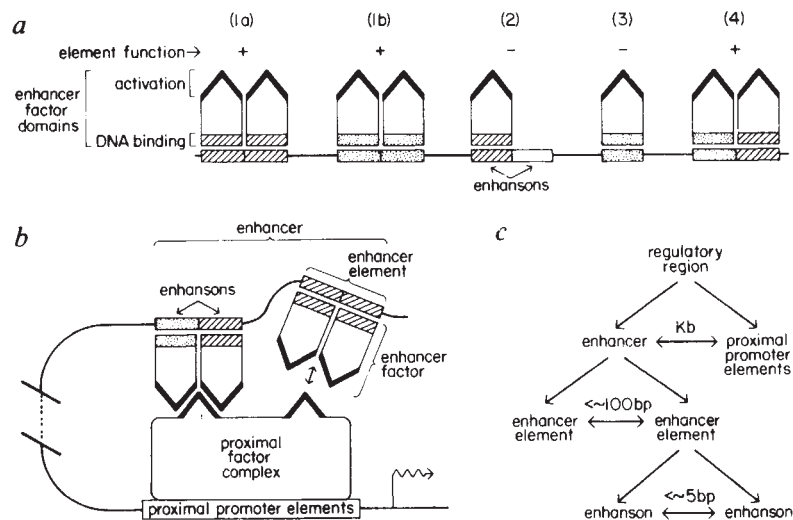
junction. We are presently testing the importance of the exact spacer sequences used to separate enhancers.

The bipartite structure of enhancer elements suggests that two *trans*-acting factors bind to adjacent enhancers to form a complex that can enhance transcription in combination with another enhancer element complex. Figure 6*a* shows a hypothetical series of productive (1*a*, 1*b* and 4) and nonproductive (2 and 3) complexes between enhancer factors and enhancers. In the productive complexes, pairs of identical (1*a* and 1*b*) or different (4) factors are shown bound to enhancers as might occur with the coreA/coreA and sphII/sphII elements (complexes 1*a* and 1*b*) or the chimaeric element sphII/coreA (complex 4). The juxtaposition of these homologous or heterologous factors hypothetically creates an activation domain (**) that is functionally equivalent between different enhancer element complexes because enhancer elements (and thus factors) can functionally compensate for one another²⁰. Complexes 2 and 3 represent solitary, nonproductive interactions between enhancers and enhancer factors. Complex 2 is inactive because the transcription factor specific for the unoccupied enhancer is absent. This could represent a cell-type-specific element and may explain the presence of motifs and *trans*-acting-factor binding sites common to many enhancers (for example, core) in cell-type-specific enhancers. Complex 3 represents a solitary enhancer which could be activated by duplication (as in complexes 1*a* and 1*b*) or rearrangement to form a chimaeric element (as in complex 4). These hypothetical non-productive complexes show how enhancer motifs could serve as binding sites for *trans*-acting factors but not display activity *in vivo*.

There are a number of reported cases of pairs of bound factors that may serve as examples of functional enhancer element complexes. With respect to the SV40 enhancer, Davidson *et al.*⁷ have presented evidence, from detailed DNaseI and dimethyl sulphate protection experiments, that HeLa cell nuclear extracts contain an sphI and sphII motif binding factor that interacts more tightly with the sphII motif. This difference in affinity is consistent with the weak activity of the sphI/sphI element compared to the more active sphII/sphII element in HeLa cells (see Fig. 4). In a separate case, the purified factors AP-2 and AP-3 (refs 10, 17, 18) bind independently to separate 'halves' of the C element¹⁷ and thus may represent different factors that cooperate to form an active element (as in complex 4, Fig. 6*a*). But, it is possible that not all enhancer elements are bipartite. Rather some enhancer elements may contain either single or multiple enhancers. For example, some factors may carry fully functional 'activation' domains. A candidate *trans*-activator in this class is AP-1 because multiple copies of a 10-bp segment of the SV40 enhancer designed to contain only the AP-1 binding site can enhance transcription¹⁰.

The relatively strict spacing requirements between paired enhancers (see Fig. 5*a*) can be explained by a need for close protein-protein contacts to form a functional enhancer element complex. It is not clear, however, how enhancer elements display less strict spacing constraints and little if any stereospecific constraints (Fig. 5*b* and ref. 5). Bienz and Pelham²⁶ have shown that a single copy of a heat-shock-response enhancer element can activate transcription when positioned near the transcriptional start site but that two heat-shock enhancer elements are required for enhancer function; that is, transcriptional activation at a distance. This result suggests that the requirement for two enhancer elements is specific to the ability to activate transcription at a distance. One model for the mechanism of enhancer function is that the proteins bound to the enhancer elements in turn associate, by DNA looping^{27,28}, with one or more proteins in a transcriptional complex bound to proximal promoter elements. This arrangement is shown schematically in Fig. 6*b*. Within this framework two enhancer elements may be required for enhancer function if the association between enhancer elements and proximal elements is broken and reformed during transcriptional initiation. The long-range enhancer/promoter

Fig. 6 Enhancer, enhancer element, and enhanson organization. *a*, Putative functional and non-functional complexes between *trans*-acting factors and enhansons. Enhansons of different sequence are shown as differently shaded boxes along a DNA molecule (line). *Trans*-acting factors are shown above with (i) shaded portions indicating DNA binding domains that correspond to a specific enhanson sequence and (ii) blackened ^ shapes indicating a half-activation domain (2 and 3) that upon association with a second enhanson binding factor creates a functional activation domain (^) (1a, b, and 4). See text for further details. *b*, Illustration of two functional enhancer element-factor complexes interacting with a proximal factor complex bound to proximal promoter elements. The structure of the enhancer element complexes is as explained in *a*. The bidirectional arrow indicates that interactions between factors bound to enhancer and proximal elements may break and reunite at alternating times during the transcriptional process. *c*, Three levels of binary organization within a regulatory region can be functionally distinguished by their differential response to spacing (see text).



interaction could be maintained by breakage and reassociation of the two separate enhancer element complexes at alternating times (see Fig. 6b).

Implications of a binary regulatory structure

The structure of enhancers described here emphasizes multiple levels of binary organization in which each level is differentiated from the other by the flexibility in spacing between its component pair of subunits. Figure 6c shows a regulatory region divided into three levels of binary organization with each level exhibiting increased sensitivity to spacing between its components. At the first level a regulatory region is divided into enhancers and proximal promoter elements (for example, CCAAT and TATA boxes); these two sets of control elements are capable of activating transcription when separated by distances as large as 10 Kb²⁹. The enhancer is in turn composed of elements that can enhance transcription when separated by roughly up to 100 bp, whereas a pair of enhansons must be in close proximity to create a functional enhancer element. The β -globin expression construct with just two SV40 enhancer B elements used in Fig. 5b exemplifies this regulatory structure in its simplest form. The enhancer is separated by over 2 Kb from the β -globin upstream promoter region and is composed of just two B elements each of which contains a pair of sph enhansons.

An enhancer structure that contains multiple levels of binary organization permits a greater degree of transcriptional regulation with a limited number of transcription factors. For example, the coreA/coreA and sphII/sphII enhancers probably respond independently to two different enhanson-specific factors whereas the sphII/coreA element is probably dependent on the

activity of both these factors. In the simplest case, the combination of two factors A and B and a single level of binary organization can generate four different regulatory combinations (AA, AB, BA and BB), but with two levels of binary organization sixteen different combinations (that is, AA-AA, AA-AB, AA-BA, etcetera) are possible. Some of these may be inactive whereas others may display different patterns of cell-specific or temporal regulation. These complicated combinatorial schemes are reminiscent of the complex appearance of the SV40 enhancer.

In addition to providing a greater degree of flexibility in transcriptional regulation, the increased complexity generated by two levels of binary organization may serve to buffer regulatory regions from the spontaneous appearance of new enhancers during evolution. Unlike proximal elements, enhancers can activate transcription over very large distances from the transcriptional start site. This high degree of positional freedom creates a much larger target size within which mutations can create new enhancers. Thus if enhancers were structurally simple, they might evolve frequently, creating chaotic patterns of gene expression.

We thank M. Tanaka for helpful discussions; M. Goodwin and R. Duffy for typing the manuscript; J. Duffy, D. Green and M. Ockler for artwork and photography; M. Iacovacci, K. Johnson and M. Zoller for synthetic oligonucleotides; V. Bautoch, M. Gilman, T. Grodzicker, D. Hanahan, N. Hernandez, R. Roberts, B. Stillman, M. Tanaka and J. D. Watson for critical readings of the manuscript; and L. Levin and C. Ales for computer assistance and discussion. W.H. is a Rita Allen Foundation Scholar. This work was supported by a grant from the National Cancer Institute.

Received 1 February; accepted 5 April 1988.

1. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299-308 (1981).
2. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047-6068 (1981).
3. Weiher, H., König, M. & Gruss, P. *Science* **219**, 626-631 (1983).
4. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6453-6457 (1982).
5. Zenke, M. *et al. EMBO J.* **5**, 387-397 (1986).
6. Wildeman, A. G. *et al. Molec. cell. Biol.* **6**, 2098-2105 (1986).
7. Davidson, I. *et al. Nature* **323**, 544-548 (1986).
8. Bohmann, D. *et al. Nature* **325**, 268-272 (1987).
9. Lee, W., Haslinger, A., Karin, M. & Tjian, R. *Nature* **325**, 368-372 (1987).
10. Chiu, R., Imagawa, M., Imbra, R. J., Bockoven, J. R. & Karin, M. *Nature* **329**, 648-651 (1987).
11. Xiao, J. H. *et al. EMBO J.* **6**, 3005-3013 (1987).
12. Xiao, J. H. *et al. Genes Dev.* **1**, 794-807 (1987).
13. Rosales, R. *EMBO J.* **6**, 3015-3025 (1987).
14. Johnson, P. F., Landschulz, W. H., Graves, B. J. & McKnight, S. L. *Genes Dev.* **1**, 133-146 (1987).

15. Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741-752 (1987).
16. Angel, P. *et al. Cell* **49**, 729-739 (1987).
17. Mitchell, P. J., Wang, C. & Tjian, R. *Cell* **50**, 847-861 (1987).
18. Imagawa, M., Chiu, R. & Karin, M. *Cell* **51**, 251-260 (1987).
19. Sturm, R., Baumruker, T., Franza, Jr., R. B. & Herr, W. *Genes Dev.* **1**, 1147-1160 (1987).
20. Herr, W. & Clarke, J. *Cell* **45**, 461-470 (1986).
21. Herr, W. & Gluzman, Y. *Nature* **313**, 711-714 (1985).
22. Clarke, J. & Herr, W. *J. Virol.* **61**, 3536-3542 (1987).
23. Ondek, B., Shepard, A. & Herr, W. *EMBO J.* **6**, 1017-1025 (1987).
24. Schirm, S., Jiricny, J. & Schaffner, W. *Genes Dev.* **1**, 65-74 (1987).
25. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71-74 (1984).
26. Bienz, M. & Pelham, H. R. B. *Cell* **45**, 753-760 (1986).
27. Dunn, T. M., Hahn, S., Ogden, S. & Schlieff, R. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017-5020 (1984).
28. Ptashne, M. *Nature* **322**, 697-701 (1986).
29. Pinkert, C., Ornitz, D., Brinster, R. & Palmiter, R. *Genes Dev.* **1**, 268-276 (1987).
30. Lusky, M., Berg, L., Weiher, H. & Botchan, M. *Molec. cell Biol.* **3**, 1108-1122 (1983).

Is the 'great attractor' a loop of cosmic string?

Y. Hoffman & W. H. Zurek

Theoretical Astrophysics, T-6, MS-B288, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Recent measurements of galaxy velocities suggest that the observed large-scale streaming^{1,2} may be attributed to a massive ($M \approx 10^{16} M_\odot$) 'attractor' located at a distance $D \approx 45 h^{-1} \text{ Mpc}$ (ref. 3; h = Hubble's constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$). We explore the idea that the streaming was induced by a large, moving loop of cosmic string. A stationary loop induces a velocity field that falls off as r^{-1} , where r is the distance from the loop. This is somewhat modified by the motion of the loop, but the r^{-1} profile still persists in much of the wake of the string. The standard gaussian δ -field (density perturbation) inflationary models of cold or hot dark matter predict, on the other hand, a velocity that should fall off as r^{-3} away from the density peak. Extension of this model to the Local Supercluster (LSC) allows one to understand its Virgocentric velocity field of r^{-1} .

The possibility that the large-scale streaming^{1,2} may be induced by cosmic string loops has been proposed by several groups⁴⁻⁶. Although their results differ in detail, their qualitative conclusions are substantially in agreement: cosmic strings fare somewhat better than the standard gaussian density perturbation (δ) field but for a universe dominated by cold dark matter (CDM), the predicted velocities are smaller than observed. The hot dark matter (HDM) case, on the other hand, yields larger velocities which are almost in agreement with the observations. All of these investigations adopt a 'Copernican' viewpoint, that we are in a random place in a universe with a statistical distribution of cosmic string loops. Consequently, these studies concentrate on calculating the probability and rms values of streaming velocities.

Motivated by the observation³ that the streaming motion is best fitted by a flow towards a single attractor, obeying approximately an r^{-1} infall, we study the velocity field induced by a single string loop. Such an infall onto a peak of density does not arise naturally in the standard gaussian CDM or HDM models, which have a Zeldovich-Harrison power spectrum on such large scales. The model we propose is of an infall seeded by a single loop embedded in a flat Einstein-de Sitter universe. The model is tested by asking (1) what are the amplitude and structure of the induced field and, (2) what is the probability of finding such a loop at the observed distance?

Let us first calculate the dynamics of the structure which has been seeded by a single loop. We assume the loop mass distribution resembles a thin spherical shell of radius r_s (refs 4, 7, 8), and allow it to be laid down with some initial peculiar velocity, v_i . To understand the basic scaling of the induced velocity field we shall consider first the case of a stationary loop (which has been derived in refs 4 and 5). Following Bertschinger⁹, we study the dynamics in the framework of the Zeldovich¹⁰ formalism, relating the eulerian and lagrangian coordinates $\vec{r}$ and $\vec{q}$ by

$$\vec{r} = a[\vec{q} + \vec{\psi}(\vec{q}, t)] \quad (1)$$

Here co-moving displacement $\vec{\psi}(\vec{q}, t)$ is the solution of the inhomogeneous equation:

$$\left[\frac{\partial^2}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial}{\partial t} + 3 \frac{\ddot{a}}{a} \right] \vec{\psi} = \begin{cases} \frac{-GM_s(\vec{q} - \vec{q}_{\text{loop}})}{a^3 |\vec{q} - \vec{q}_{\text{loop}}|^3} & \text{if } \vec{q} \text{ is outside the loop at } t; \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This equation is solved subject to $\vec{\psi} = \partial \vec{\psi} / \partial t = 0$ at t_s , where t_s

is the time when the loop was laid down, which we assume to be in the matter dominated era. In these expressions, a is the scale factor (normalized by $a(t_s) = 1$), M_s is the mass of the loop and $\vec{q}_{\text{loop}}$ is its position, which in the case of a stationary loop is taken to be at $\vec{q} = 0$. The velocity field induced by a point-like stationary mass decreases as r^{-2} (refs 4, 5), and therefore points which at t_s were exterior to the loop exhibit the same profile. However, interior points (at t_s) start to experience gravitational instability only after the Hubble flow takes them outside the loop at the time corresponding to the scale factor $a_i = r_s/q$.

This modifies the velocity profile to

$$\vec{v}(\vec{q}, a) = -\frac{3GM_s t_s (1 + Z_s)^{3/2}}{5r_s^2} \vec{r} \quad (3)$$

where r_s is the physical radius of the loop and G is the gravitational constant. For a cosmic string of tension μ we have $M_s = \beta \mu r_s$, where β is a constant which depends on the loop trajectory. Recently, two independent numerical simulations put it in the range of 15 to 25 (N. Turok and D. Bennett, personal communication), considerably above the initial estimate⁴ of $\beta \approx 9$. The loop radius, r_s , is a fraction ε of the horizon length at t_s , $r_s = \varepsilon c t_s$, where c is the speed of light. This velocity field is divergence-free for the exterior points which leads to a vanishing (to first order) δ -field at that regime⁹. For interior points, on the other hand, the r^{-1} velocity field corresponds to a δ -profile which decreases as r^{-2} . A spherical shell of dark matter which is concentric with the loop and has a lagrangian radius $q = r_s$ at redshift Z_s at which the loop is laid down would experience a present day infall with peculiar velocity

$$v_{500} = 3.6 \times 10^{-2} \frac{\beta_{20} \mu_{-6}}{\varepsilon_{0.2}} (1 + Z_s)^{1/2} \quad (4)$$

where v_{500} is in units of 500 km s^{-1} , $\varepsilon_{0.2} = \varepsilon/0.2$, $\beta_{20} = \beta/20$ and $\mu_{-6} = 10^6 [G\mu/c^2]$. The present epoch radius of this shell is

$$R = 14.5 \left[\frac{\beta_{20} \mu_{-6}}{v_{500}} \right] h^{-1} \text{ Mpc} \quad (5)$$

The epoch at which the loop was laid down is constrained by equations (4) and (5). Yet the r^{-1} velocity field might extend beyond the Local Group which implies that $r_s \geq q(45 h^{-1} \text{ Mpc})$, and therefore that constraint constitutes an upper bound on Z_s . However, note that the basic R versus v_{500} relation is independent of Z_s .

The case of a moving loop is necessarily axisymmetric around its direction of motion (defined as the z -axis) and we calculate its velocity and δ -field in the x - z plane. The general solution for ψ depends on when and how many times the trajectories of dark matter particles cross the loop. We look for all the positive a_i roots of the equation

$$q_x^2 + (q_z - q_{\text{loop}}(a_i))^2 = \left(\frac{r_s}{a_i} \right)^2 \quad (6)$$

Here a_i are the scale factors associated with the time instants at which the points cross the perimeter of the loop and $q_{\text{loop}}(a)$ is given by⁹

$$q_{\text{loop}}(a) = D_i(1 - a^{-1/2}) \quad (7)$$

Here $D_i = 3v_i t_s$ and v_i is the initial string velocity. Note that equation (6) is valid only in the linear regime where the displacement field $\vec{\psi}$ is negligible. There are four possible roots for $a_i^{1/2}$. The case of no real roots larger than unity corresponds to points that are always outside the loop (defined as exterior points), and, for them, a_i is set to be 1. Most of the points that were interior at t_s have only one root larger than unity, but a small subset of points show a more complex behaviour. For example, some have two roots larger than unity, corresponding to points that started outside of the loop, moved in and then outside

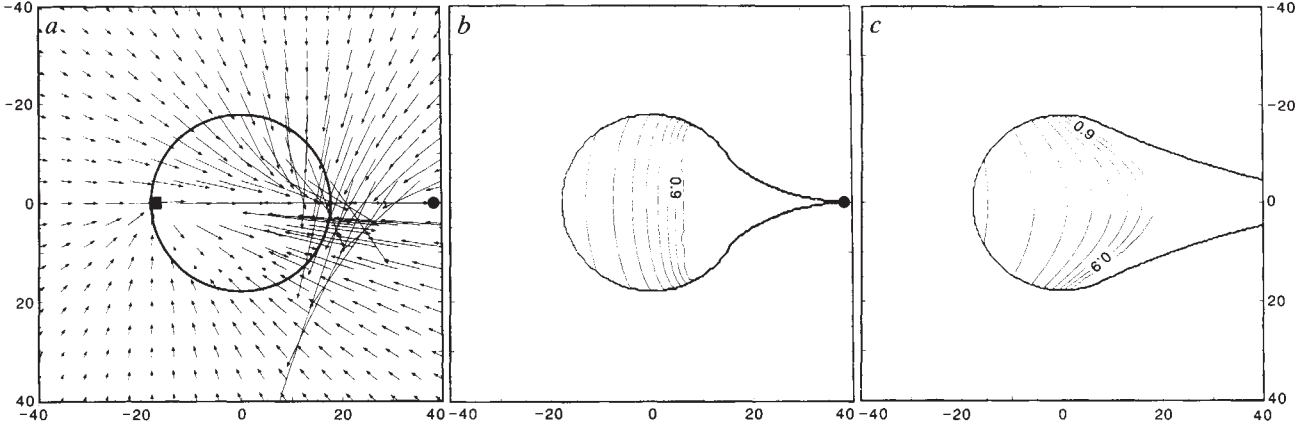


Fig. 1 The peculiar velocity and overdensity δ induced by a moving loop of cosmic string. *a*, Vector plot of v where the upper half of the figure gives the actual velocity field. The bottom part shows the line-of-sight peculiar velocity seen by an observer located at the point marked by the square. Both velocities are with respect to the microwave background. The present location of the string is marked by the dot and the initial (co-moving) radius of the loop is given by the circle (thick line). *b*, Isodensity plot of the δ -field, where the overdensity contours are going from 0.1 to 1.0 in steps of 0.1 and the three lowest δ contours overlap. Points that have always been outside of the loop have a vanishing δ . As η increases above ~ 0.2 the isodensity contours assume a trumpet-like shape (*c*). The string shown in the figure was laid down at redshift $Z_s = 500$, with initial mass $M_s = \beta\mu\epsilon ct_s$, where $(\beta/20)(10^6 G\mu/c^2) = 2.25$ and $\epsilon/0.2 = 1$, and its initial peculiar velocity is $\eta = v_i/c = 0.15$, except *c*, where $\eta = 0.3$. Distances are in units of h^{-1} Mpc and the magnitude of the velocity vectors is given by their Hubble length.

again. The displacement and velocity of the points with only one root $a_i \geq 1$ is given by

$$\vec{\psi} = -\psi_0 a(t) \{ [g_1(\vec{q}, a) - g_1(\vec{q}, a_i)] \vec{e}_x + [g_2(\vec{q}, a) - g_2(\vec{q}, a_i)] \vec{e}_z \} \quad (8)$$

$$v_z = -V_0 \left[g_2(\vec{q}, a) - g_2(\vec{q}, a_i) + \frac{D_z^2(q_z - q_{loop})}{2R^3(\vec{q}, a)a} \right] \sqrt{1+Z_s} \quad (9)$$

$$v_x = -V_0 \left[g_1(\vec{q}, a) - g_1(\vec{q}, a_i) + \frac{D_x^2 q_x}{2R^3(\vec{q}, a)a} \right] \sqrt{1+Z_s} \quad (10)$$

where

$$\psi_0 = \frac{1}{5} \frac{\beta\epsilon G\mu}{\eta^2 c^2} ct_s (1+Z_s)^{-3/2},$$

$$V_0 = \frac{2}{15} \frac{\beta\epsilon G\mu}{\eta^2 c^2} c,$$

$$R(\vec{q}, a) = [q_x^2 + (q_z - q_{loop})^2]^{1/2},$$

while $\eta = v_i/c$ and the functions g_1 and g_2 of $\vec{q}$ and a are defined in ref. 9. The cases of points with more than one $a_i \geq 1$ root are solved by combining solutions of the form given by equations (8)–(10) with solutions of the source-free equation for $\vec{\psi}$, with the appropriate initial conditions. The density field is then given by $\delta(\vec{q}, t) = -(\vec{\nabla} \cdot \vec{\psi})$.

The velocity field is presented in the form of a vector plot in Fig. 1, where half the plot corresponds to actual peculiar velocities and the other half gives the line-of-sight component relative to an observer located at $q_z = -r_s$ (in both cases velocities are relative to the reference frame at rest with respect to the microwave background). Figure 1 also shows the δ -field in the form of a contour plot. To see the effect of the loop velocity we calculated the velocity field of a stationary loop and a one that starts with $\eta \approx 0.15$, both laid down at $Z_s = 500$; the string parameters are $\beta_{20}\mu_{-6} = 2.25$ (see Fig. 2). Figure 1 shows the infall to be centred roughly on the region where δ attains its maximal value, which is located in the wake behind, and not coinciding with, the present location of the loop. Its exact position depends on η , and for $\eta = 0.1$ it is at $q_z \approx \frac{2}{3}r_s$. Let us first consider the effect v_i has on the velocity profile along the z -axis. At $q_z \leq -r_s$ the profile becomes steeper than r^{-2} , goes to r^{-1} at $q_z \approx -r_s$ until it reaches a maximal value, then starts to decline and changes sign with a profile steeper than r^{-1} which

goes eventually into r^{-2} (here r is the distance from the 'centre' of attraction). For off-axis points the values of v_r (radial component of v relative to this 'center') are in between the values corresponding to the up- and down-stream points along the z -axis, at a given r . As η increases, the region over which the r^{-1} profile prevails first increases, but then starts to decrease (at $\eta \approx 0.15$) until it disappears at $\eta \approx 0.5$. The isodensity contours are quasi-spherical for $\eta \leq 0.2$. This changes to a trumpet-like shape as η increases, in the sense that the δ dependence on x at a constant z is not monotonic. It has a local minimum on the z -axis, and as $|x|$ increases, δ increases to a maximum and then falls to zero outside the loop. Note that Figs 1 and 2 give the formal solution of the linear theory, which does not apply at that part of the wake close to the loop where ψ becomes non-linear. A nonlinear calculation would eliminate the appearance of the discontinuity in δ at the boundary of the region containing the interior points.

The basic scaling of the model is given by equations (4) and (5), which apply for a stationary loop. Applying this simplified model to the case of the observed 'attractor' we find infall velocity of $v_{500} \approx 1$ at $r = 45 h^{-1}$ Mpc for $\beta_{20} \mu_{-6} \approx 2.7$ (we assume here $r \approx R$) which implies $Z_s \sim 105$. The string tension μ is a universal parameter, and is estimated to be around $\mu_{-6} \approx 2$ in the case of a CDM dominated universe¹¹, and the string tension can be as large as $\mu_{-6} \approx 5$ in the HDM case⁴. A CDM model predicts the scale R , or equivalently streaming velocity v , to be of the order of the required minimum. Moreover, for a HDM model with $\mu_{-6} = 5$ and $\beta_{20} = 1$, the scale R and the streaming velocity are well within the range of the observed values. Introducing a non-vanishing v_i helps the string model in reproducing the observed infall, in the sense that a given velocity at a given distance can be obtained for a much higher value of Z_s , that is by a loop much smaller in size and mass. Infall velocities of about 450 km s^{-1} in the wake of a loop with $Z_s = 500$, $\eta = 0.15$ and $\beta_{20}\mu_{-6} = 2.25$, are now expected at $40 h^{-1}$ Mpc away from the attractor. Note that our estimates are based on co-moving distances. Taking into account the displacement ψ would decrease r by $\sim 20\%$.

A note of caution concerning the estimated string parameters should be sounded here. The string tension, μ , can be estimated from the masses of Abell clusters. Turok and Brandenberger¹¹ inferred $\mu_{-6} \approx 2$ for CDM, assuming that the Abell-cluster seeding loops were laid down at the end of the radiation dominated era. On the other hand, in fitting the richness distribution of

Abell clusters, one finds an estimate of the product $\mu\epsilon \approx 2$ (for $h = 0.5$)⁸. For $\epsilon \approx 0.2$ this would imply a rather large μ_{-6} , but both of these estimates depend on the values of β and other parameters of the string network as well as on the parameters of the Abell cluster mass distribution, which leads to further uncertainty by a factor ≈ 2 . These uncertainties presently give the CDM model enough room to accommodate the observed velocities.

Attention has been paid so far to the magnitude of the velocity, but equally important is the dependence of v on r . The observed approximate r^{-1} dependence is inconsistent with the great attractor being centred on a peak of density in both CDM and HDM gaussian δ -field inflationary cosmogonies. In these models the (ensemble) mean density distribution $\delta_0(r)$ around very high local δ -maxima is given by¹²⁻¹⁵

$$\delta_0(r) = \delta_0 \frac{\xi(r)}{\sigma^2}. \quad (11)$$

Here δ_0 is the value of δ at the peak, $\xi(r)$ is the autocorrelation function and $\sigma^2 = \xi(0)$. For lower peaks $\delta_0(r)$ is steeper than the r dependence of equation (6) (due to the curvature dependence of the correlation matrix). For a scale-free gaussian δ -field defined by its power spectrum $p(k) \propto k^n$, the correlation function follows $\xi(r) \propto r^{-(n+3)}$. Assuming that the 'great attractor' has been seeded by a local δ -maximum this implies an infall velocity onto the peak that goes as $r^{-(n+2)}$. Both the CDM and HDM models have a power spectrum which is close to $p(k) \propto k$ on the scale of a few tens of Mpc. This yields a mean infall profile of r^{-3} , which makes the observed profile a very unlikely outcome of such models. Note that the velocity field in the Local Supercluster (LSC) also shows an r^{-1} dependence^{16,17}, again in disagreement with the CDM and HDM models, which predict a velocity profile somewhat steeper than r^{-2} on that scale. The string model explains this velocity field naturally, but to accommodate the velocity amplitudes of the LSC and 'cosmic attractor' one would have to rely on the intrinsic spread in some of the loop parameters (for example β and ϵ), but not on a different Z_s , as the basic radius-velocity relation (5) is independent of Z_s .

How likely is it for a loop of the size required to seed the 'cosmic attractor' to be deposited within an acceptable distance away from us? To answer this question we shall calculate the radius, R_s , of a sphere expected to contain one loop of size r_s . The number density of the loops of size r_s , in the matter dominated era, is given by¹¹:

$$dn(r_s) = \frac{\nu dr_s}{r_s^2 (ct_s)^2} \quad (12)$$

Here ν is a dimensionless constant whose still uncertain value is estimated to be no less than $\nu \approx 0.03$ for loops laid down in the matter dominated era⁴. Setting $dr_s \approx r_s$, we conclude that $\sim \nu$ loops of sizes between $0.5 r_s$ and $1.5 r_s$ would be deposited in a volume $(ct_s)^2 r_s = \epsilon (ct_s)^3$. Therefore, one loop of the radius of that order can be typically found in a sphere of radius $R_s = (3\epsilon/4\pi\nu)^{1/3} ct_0/\sqrt{1+Z_s}$.

This yields (for a stationary loop)

$$R_s = \frac{\beta_{20}\mu_{-6}}{\nu_{0.03}^{1/3} v_{500} \epsilon_{0.2}^{2/3}} 84 h^{-1} \text{ Mpc} \quad (13)$$

where $\nu_{0.03} = \nu/0.03$. For $\beta_{20}\mu_{-6} \approx 2.25$ this implies a large, but not prohibitive distance of $\sim 230 h^{-1}$ Mpc. For moving loops, R_s can be made somewhat smaller with increase of Z_s , but this is offset by the fact that only in a fraction of the volume—in the wake of the moving string—are the desired velocities obtained. We find that $\approx 1\%$ of spheres of radius $R \approx 45 h^{-1}$ would contain such a loop, and therefore our results do not contradict the previous calculations of r.m.s. velocities⁴⁻⁶.

It should be emphasized that the above argument assumes that the sphere is centred on a randomly placed point in space,

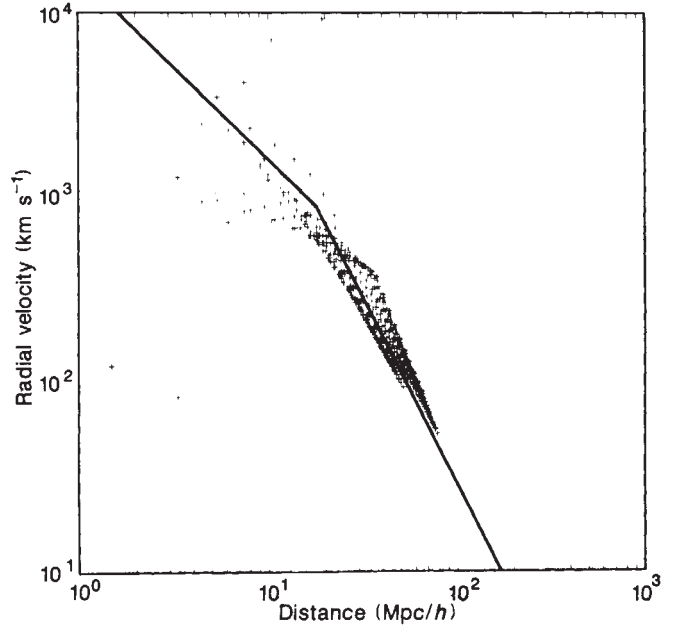


Fig. 2 A scatter-plot of the radial velocity against distance of all grid points of Fig. 1. The centre is defined as the point along the z -axis where v_z changes sign. The solid line corresponds to a stationary loop and the scatter-plot to a moving loop of $\eta = 0.15$; all other parameters are as in Fig. 1. In both cases velocity falls off as r^{-1} in a large fraction of the volume, where r is the distance from the nonlinear part of the diagram, the presumed location of the 'great attractor'. This velocity-distance relation provides the best fit to the large scale streaming data³.

not on a randomly chosen galaxy. Loops of the same size are strongly correlated¹⁸. If the same is true for loops on different scales (for loops seeding galaxies, clusters and 'attractors') the estimate of the distance to the nearest loop of radius r_s needed to seed the attractor is far too generous. Such correlation between galaxies and loops seeding large scale features could solve the Ω_0 -problem, and can be expected if loop self-intersection¹⁹ plays an important role.

What are the observational consequences of the existence of such a loop and the object seeded by it? We first consider the properties of the loop itself at the present epoch. It is moving with a peculiar velocity of $v_{\text{loop}} = \eta[(1+Z_s)/500]^{-1} 600 \text{ km s}^{-1}$, its mass is $M_s = 1.3\beta_{20}\mu_{-6}\epsilon_{0.2}[(1+Z_s)/500]^{-3/2} 10^{13} h^{-1} M_\odot$, its radius is $r_s = \epsilon_{0.2}[(1+Z_s)/500]^{-3/2} 3.58 \times 10^{-2} h^{-1} \text{ Mpc}$, and it is located today at a distance $D = 15(\eta/\epsilon_{0.2})(1+Z_s)r_s \times (1 - (1+Z_s)^{-1/2})$ from its starting point, which for the parameters used in Fig. 1a is $D \approx 38 h^{-1} \text{ Mpc}$. Thus, an observer located on the z -axis at $z = -R$ would see the loop at a distance of $\sim 56 h^{-1} \text{ Mpc}$ away, at the other side of the 'cosmic attractor'. The properties of the cluster seeded by the loop depend on its initial velocity. The densest objects are seeded by the stationary loops. In the case of a stationary loop with the parameters of Fig. 1a the mean overdensity within a sphere of radius R is only 1.5 (assuming linear evolution of δ). An inner sphere that contains a typical rich Abell cluster mass of $10^{15} M_\odot$ would collapse and virialise at about $0.6 T_0$, where T_0 is the age of the universe, which makes it a possible candidate for an Einstein-like X-ray telescope. However in the more realistic case of a loop moving with $v_{\text{loop}} \approx (0.1-0.2)c$, the central density would be smaller and the object seeded by its wake would look more like a supercluster with no excessively rich cluster at its center, not unlike the Local Supercluster. Such an object would not be prominent in the Einstein observatory survey, yet redshift surveys should find an extended supercluster centred on the observed attractor. Indeed recent surveys^{20,21} found an extended concentration of galaxies roughly centred on the attractor. Moreover, as we have already pointed out above, an object

containing the loop would be now beyond the apparent location of the attractor inferred from the velocity field. Numerical study of the nonlinear evolution of the initial density perturbation is clearly required to estimate properties of its densest parts more reliably. The direct detection of the loop or the 'cluster' seeded by it is even more difficult in the case of the observed attractor because of its low latitude.

The large-scale velocity field has been modeled here as an infall induced by a single moving string loop. The present work complements the other recent studies⁴⁻⁶ of the r.m.s. value of the streaming velocities, and is consistent with them in the following sense: for the string parameters that yield streaming motion in agreement with the observed value, the infall velocity we find is also consistent with the observations. Yet we add here another dimension to the problem and consider the spatial variation of the infall velocity field. We find the r^{-1} dependence is a natural outcome of the cosmic string scenario.

Finally, we should point out that both the finite loop size and its peculiar velocity will have to be taken into account in studies of the structure of string-seeded objects. Velocity of the loop has already proved to be important for the seeds of galaxies²², and both loop size and its peculiar velocity should play a role in determining the structure of Abell clusters⁹. Therefore, the results presented above will be needed to set up the initial conditions for numerical simulations of structure formation seeded by loops of cosmic strings.

Received 20 November 1987; accepted 17 February 1988.

1. Burstein, D. *et al.* in *Galaxies, Distances and Deviations from Universal Expansion* (eds Madore, B. F. & Tully, R. B.) 123-130 (Reidel, Dordrecht, 1986).
2. Dressler, A. *et al.* *Astrophys. J.* **313**, L37-L42 (1987).
3. Lynden-Bell, D. *et al.* *Astrophys. J.* **326**, 19-49 (1988).
4. Brandenberger, R., Kaiser, N., Shellard, E. P. S. & Turok, N. *Phys. Rev. D* **36**, 335-345 (1987).
5. Bertschinger, E. *Astrophys. J.* **324**, 5-18 (1988).
6. Van Dalen, A. & Schramm, D. N. *Astrophys. J.* **326**, 70-76 (1988).
7. Sato, H. *Prog. theor. Phys.* **75**, 1342-1350 (1986).
8. Zurek, W. H. *Astrophys. J.* **324**, 19-34 (1988).
9. Bertschinger, E. *Astrophys. J.* **316**, 489-496 (1987).
10. Zeldovich, Ya. B. *Astr. Astrophys.* **5**, 84-89 (1970).
11. Turok, N. & Brandenberger, R. *Phys. Rev. D* **33**, 2175-2181 (1986).
12. Doroshkevich, A. G. *Astrophys. J.* **320**, 320-330 (1970).
13. Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. *Astrophys. J.* **304**, 15-61 (1986).
14. Hoffman, Y. & Shaham, J. *Astrophys. J.* **297**, 16-22 (1985).
15. Hoffman, Y. *Astrophys. J.* **328** (in the press).
16. Yahil, A., Sandage, A. & Tammann, G. in *Physical Cosmology* (eds Balian, R., Audouze, J. & Schramm, D. N.) 127-159 (North-Holland, Amsterdam, 1980).
17. Aaronson, M., Huchra, J., Mould, J., Schechter, P. L. & Tully, R. B. *Astrophys. J.* **258**, 64-76 (1982).
18. Turok, N. *Phys. Rev. Lett.* **55**, 1801-1804 (1986).
19. Zurek, W. H. *Phys. Rev. Lett.* **57**, 2326-2329 (1986).
20. Dressler, A. *Astrophys. J.* (in the press).
21. Strauss, M. A. & Davis, M. in *Proc. Vatican Study Week on Large-Scale Motions in the Universe* (in the press).
22. Zurek, W. H. & Quinn, P. J. *Proc. IAU Symp.* **130** (in the press).

Hot gaseous coronae of early-type galaxies and their radio luminosity function

Gopal-Krishna* & Paul J. Wiita†

* Tata Institute of Fundamental Research, PO Box 1234, Indian Institute of Science Campus, Bangalore 560 012, India

† Department of Physics and Astronomy, Georgia State University, Atlanta, Georgia 30303, USA

Recent X-ray observations have revealed that early-type galaxies (which usually produce extended double radio sources) generally have hot gaseous haloes extending up to $\sim 10^2$ kpc^{1,2}. Moreover, much of the cosmic X-ray background radiation is probably due to a hotter, but extremely tenuous, intergalactic medium (IGM)³. We have presented⁴⁻⁷ an analytical model for the propagation of relativistic beams from galactic nuclei, in which the beams' crossing of the pressure-matched interface between the IGM and the gaseous halo, plays an important role. The hotspots at the ends of

the beams fade quickly when their advance becomes subsonic with respect to the IGM. This model has successfully predicted (for typical double radio sources) the observed⁸ current mean linear-size ($\sim 2D \approx 350$ kpc)^{4,5}, the observed⁸⁻¹¹ decrease in linear-size with cosmological redshift⁴⁻⁶ and the slope of the linear-size versus radio luminosity^{10,12-14} relation⁶. We have also been able to predict the redshift-dependence of observed numbers and radio luminosities of giant radio galaxies^{7,15}. Here, we extend this model to include the propagation of somewhat weaker beams. We show that the observed flattening of the local radio luminosity function (LRLF)¹⁶⁻²⁰ for radio luminosity $P \approx 10^{24}$ W Hz⁻¹ at 1 GHz can be explained without invoking *ad hoc* a corresponding break in the beam power function $\Phi(L_b)$, because the heads of the beams with $L_b < 10^{25}$ W Hz⁻¹ are decelerated to sonic velocity within the halo itself, which leads to a rapid decay of radio luminosity and a reduced contribution of these intrinsically weaker sources to the observed LRLF.

We adopt average, observationally determined^{1,21} parameters for the haloes. The atomic number density is $n(r) = 10^{-2} (1 + (r/a)^2)^{-0.75}$ cm⁻³ where $a = 2$ kpc and $kT_h = 1$ keV; we have argued⁴ that these properties have not changed very much since a cosmological redshift $z \approx 1$. Average parameters for the IGM from an analysis of the X-ray background² are density $n_{\text{IGM}} = 7 \times 10^{-7} (1+z)^3$ cm⁻³ and $kT_{\text{IGM}} = 18 (1+z)^2$ keV. The higher pressure of the IGM at earlier epochs implies that the pressure-matched interface between these two media, for typical isolated ellipticals, is at a radius, $R_h(z) = 170 (1+z)^{-10/3}$ kpc. Our analytical beam models are based upon the emission of relativistic fluid with relativistic bulk velocity²² into two narrow cones of opening angle θ . The velocity of advance of the head of the beam, v_h , is determined by ram-pressure balance with the current external medium; we find^{4,5} within the quasi-power-law halo, $v_h \propto r^{-0.25}$, while in the IGM, $v_h \propto r^{-1}$. For very high powered beams, even v_h is relativistic for some time⁷, slightly modifying these results. We have argued^{4,7} that the source should fade rapidly when v_h falls below, c_s , the sound speed in the external medium, or when the lifetime of the source exceeds the total period of nuclear activity, $t_N \approx 10^8$ yr.

Analytically, there are two extreme cases⁴⁻⁷ for the supersonic advance of the beam head upon crossing the halo/IGM interface. In model A, θ is assumed to remain constant across the interface, implying a temporary acceleration of the beam head because of the greatly reduced ambient density, and therefore ram pressure. In model B, the head velocity is taken to be constant across the interface, implying an abrupt flaring of the jet upon crossing into the IGM. Numerical simulations⁶ for beam propagation across a halo/IGM interface support the validity of our analytical work (P.W., Rosen and Norman, in preparation) and indicate that the reality lies roughly midway between these extremes. The smaller value of R_h along with the higher value of the sound speed of the IGM at earlier epochs explains why we can reproduce the observed^{10,11,14} median size for double radio sources $2D \approx 350$ kpc $(1+z)^{-3}$. The numerical simulations⁶ also yield a good match to the observed relation $D \propto P^{0.3}$ (refs. 10, 12-14). Further, these simulations indicate that the slopes of the D vs. z and D vs. P relations are rather insensitive to the uncertain values of t_N , $n_{\text{IGM}}(z=0)$ and T_{IGM} ; still, additional work may enable us to constrain these parameters.

To estimate the observed radio luminosity, $P(t)$, for a beam of total (mostly kinetic) beam power, L_b , we start from observations that indicate that the initial radio efficiency, ϵ , is typically ~ 0.1 (so $P_0 \equiv P(t=0) = 2\epsilon L_b$)²³⁻²⁵. We also include the gradual reduction in radio emission from inverse Compton scattering of the synchrotron-emitting relativistic electrons on the microwave background photons^{26,27}; for a given external density, we compute the speed of advance of a hotspot, and using ram pressure confinement and assuming energy equipartition between the relativistic electrons and magnetic fields we can estimate the hotspot energy density and magnetic fields.

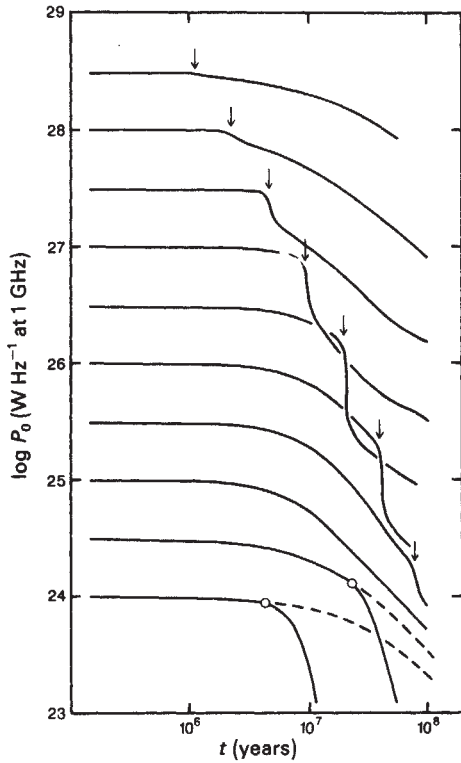


Fig. 1 The computed time-evolution of radio luminosity, P , at $z = 0.1$, for the complete composite model (see text), for a range in radio luminosity P_0 taken as a fraction $\epsilon (=0.1)$ of the total beam power $2L_b$. Arrows denote the time when the beam crosses the halo/IGM interface. The point at which the velocity of the beam head becomes sonic and subsequent rapid decay sets in is marked with a circle on each curve; the solid curves include all loss mechanisms, while the dashed extensions illustrate the putative evolution ignoring post-sonic decay within the halo.

Observations indicate²⁸ that typical lobe energy densities are roughly 10% of those in the hotspots, so we are able to estimate the magnetic energy density in the lobes as $u_b \approx 0.1 (9/7) \rho_{ext} v_h^2$. But u_b cannot fall below a value $(9/14)P_{ext}$ determined by the static pressure balance with the ambient medium⁷. (If the lobes are slowly expanding their energy densities will be only slightly above this minimum value and will not appreciably change our results). Because the flux from extended sources is dominated by emission from the lobes, we then obtain, for a given beam power and redshift, $P(t) = P_0 / (1 + u_{ph}(z)/u_b)$, with $u_{ph}(z) \propto (1+z)^4$ being the energy density of the microwave background.

Thus at a small redshift of ~ 0.1 , the typical radio luminosity, P , of a source (emitted at its mean age, $t_N/2$) is estimated⁷ to be a factor η times its initial luminosity, P_0 , where η is $\sim 30\%$ for powerful sources ($P \approx 10^{28} \text{ W Hz}^{-1}$ at 1 GHz, taking Hubble's constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$), but only $\sim 3\%$ for moderate sources with $P \approx 10^{25} \text{ W Hz}^{-1}$ (Fig. 1). We refine our model by introducing a dependence of the beam opening angle on the beam power, $L_b (=P_0/2\epsilon)$, with $\epsilon = 0.1$, which allows for the $P-\theta$ correlation observed²⁹ for the jets of classical double sources: the form used is $\theta(\text{radian}) = 0.02 + 0.03[29 - \log(2\epsilon L_b)] = 0.02 + 0.03[29 - \log(P/\eta)]$. Reinforced by the numerical simulations⁶ and our success in obtaining the properties of giant radio galaxies by using a composite model, the geometric mean of the analytical models A and B⁷, we adopt the same model to describe the time evolution of radio output, $P(t)$.

Figure 1 displays the model predictions for $P(t)$ for a wide range of P_0 evaluated at $z = 0.1$. For higher P_0 , the decrease of P with time is due solely to inverse Compton losses, which become more pronounced at the interface due to the sudden drop in the ram pressure and the concomitant faster fall in u_b . At $P_0 < 10^{25} \text{ W Hz}^{-1}$, however, another decay mechanism

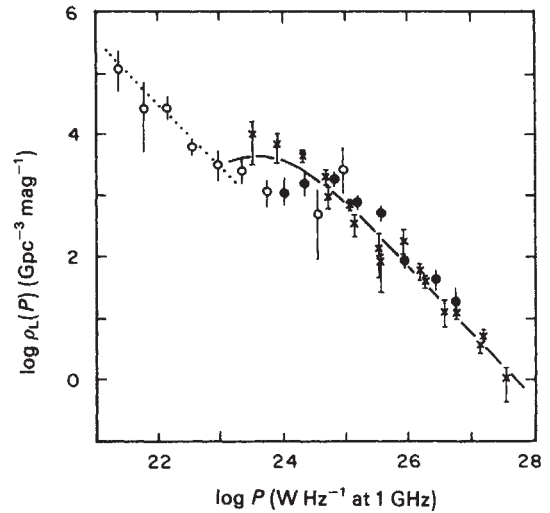


Fig. 2 The solid curve is the computed local radio luminosity function $\rho_L(P)$ at $z = 0.1$ for our model galaxy population, (E+SO), assuming a constant logarithmic slope of -1 for the beam power function $\Phi(L_b)$. The data points are the most recent determinations of $\rho_L(P)$ for early-type galaxies, translated from 843 MHz (open circles, ref. 20) and from 1.4 GHz (filled circles, ref. 17; crosses, ref. 18) assuming a spectral index $\alpha = -0.7$. The dotted line denotes the low-luminosity branch of $\rho_L(P)$ established in ref. 20. A Hubble constant of $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ has been assumed.

becomes important, as a result of the head velocity decreasing to the sonic value within the halo. From this point onwards, the relativistic beam plasma would probably advance as a quasi-ordered string of bubbles, rather than a continuous, well-collimated flow^{30,31}, rising due to buoyancy with the sonic speed, $c_s(\text{halo})$ and maintaining equilibrium with the thermal pressure, p , of the halo. For $(r/a)^2 \gg 1$, the halo pressure $p \approx 7.8 \times 10^{21} r^{-1.5}$ (in cgs units), yielding $\dot{p} \approx -3.8 \times 10^{-15} c_s(\text{halo}) p^{5/3}$. Using the minimum energy criterion, $P \approx 2.6 \times 10^{-7} U p^{3/4}$, where the pressure p within a bubble, matched^{32,33} by the halo pressure, is equal to its internal energy, U , divided by three times its volume, V . Finally, ignoring losses other than those due to expansion and synchrotron radiation, we have $dU = -p dV - P dt$. It can now be shown that during the buoyant drift starting from the sonic point, $P \propto p^2 \propto r^{-3} \propto t^{-3}$ (Fig. 1).

We now argue that a power-law for the beam power function $\Phi(L_b = P_0/2\epsilon) \propto P_0^{-1}$, as inferred from the high luminosity part of the LRLF, $\rho_L(P)$, could extend to lower beam powers and still be in accord with the observed flattening of the LRLF at $P \approx 10^{24} \text{ W Hz}^{-1}$. Only sources with $P_0 > P$ can contribute to $\rho_L(P) dP$; each bin of ΔP_0 would contribute to $\rho_L(P) dP$ for a time interval $\Delta t(P_0)$, which is estimated from Fig. 1 and integrated to yield

$$\rho_L(P) = \int_P^\infty \Delta t(P_0) \Phi(P_0) dP_0 / \int_P^\infty \Delta t(P_0) dP_0.$$

The computed $\rho_L(P)$ is compared in Fig. 2 with the most recent determinations of the LRLF. The good agreement supports our picture that the flattening of the LRLF around $P \approx 10^{24} \text{ W Hz}^{-1}$ is due to the enhanced radio decay of sources which turn on with moderate luminosities. As their beams are incapable of ploughing through the entire halo medium supersonically, lobes of such radio sources would fade out rapidly as well as lose their edge-brightened morphology, and thereby contribute less to the observed number of radio galaxies, despite their large numbers. It should be noted that our model employs the best average observationally based values for the parameters characterising the media and the beams, the latter being available²⁹ essentially for sources of only high to moderate radio luminosities. At lower luminosities another steep branch²⁰ begins to dominate the LRLF, whose properties are little understood

at present but which may appreciably contribute to samples selected at luminosities up to $P \approx 10^{24} \text{ WHz}^{-1}$ (Fig. 2). The extent of this contamination needs to be assessed in view of some recent indications^{34,35} of statistical differences between the optical properties of sources occurring on the two sides of the break in LRLF, particularly in respect of optical morphology of the host (early-type) galaxy and the degree of galaxy clustering around it. The optical absolute magnitudes, nevertheless, seem to remain confined to a fairly narrow range^{18,20}, which is consistent with our assumption that parent galaxies of sources covering several decades in radio luminosity are capable of sustaining similar hot gaseous haloes¹. To examine the possible role of the cluster environment we plan to extend our models to incorporate the effects of an intracluster medium between the halo and IGM.

We thank L. Saripalli, C. R. Subrahmanya, V. Borkar and V. K. Kapahi for helpful conversations, and A. Rosen for reading the manuscript. P.J.W. acknowledges receiving support from the Foreign Currency Research Program of the Smithsonian Institution and from a Georgia State University research grant.

Received 21 September 1987; accepted 17 February 1988.

1. Forman, W., Jones, C. & Tucker, W. *Astrophys. J.* **293**, 102–119 (1985).
2. Sarazin, C. L. *Rev. Mod. Phys.* **58**, 1–115 (1986).
3. Guilbert, P. W. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **220**, 439–451 (1986).
4. Gopal-Krishna & Wiita, P. J. *Mon. Not. R. astr. Soc.* **226**, 531–542 (1987).

5. Wiita, P. J. & Gopal-Krishna in *13th Texas Symposium on Relativistic Astrophysics* (ed. Ulmer, M. P.) 355–357 (World Scientific, Singapore, 1987).
6. Rosen, A. & Wiita, P. J. *Astrophys. J.* (in the press).
7. Gopal-Krishna, Wiita, P. J. & Saripalli, L. *Mon. Not. R. astr. Soc.* (submitted).
8. Kapahi, V. K. *Mon. Not. R. astr. Soc.* **214**, 19–23 (1985).
9. Eales, S. A. *Mon. Not. R. astr. Soc.* **217**, 179–203 (1985).
10. Oort, M. J. A., Katgert, P., Steeman, F. W. H. & Windhorst, R. A. *Astr. Astrophys.* **179**, 41–59 (1987).
11. Singal, A. K. *Mon. Not. R. astr. Soc.* (in the press).
12. Gavazzi, G. & Perola, G. C. *Astr. Astrophys.* **66**, 407–416 (1978).
13. Ekers, R. D., Fanti, R., Lari, C. & Parma, P. *Astr. Astrophys.* **101**, 194–214 (1981).
14. Kapahi, V. K. in *Highlights of Astronomy* (ed. Swings, J. P.) Vol. 7, 371–376 (Reidel, Dordrecht, 1986).
15. Saripalli, L., Gopal-Krishna, Reich, W. & Kühr, H. *Astr. Astrophys.* **170**, 20–26 (1986).
16. Auriemma, C. et al. *Astr. Astrophys.* **57**, 41–50 (1977).
17. Meier, D. L., Ulrich, M.-H., Fanti, R., Gioia, I. & Lari, C. *Astrophys. J.* **229**, 25–38 (1979).
18. Windhorst, R. A. *Thesis. Univ. Leiden* (1984).
19. Cordey, R. A. *Mon. Not. R. astr. Soc.* **219**, 575–588 (1986).
20. Subrahmanya, C. R. & Harnett, J. I. *Mon. Not. R. astr. Soc.* **225**, 297–306 (1987).
21. Trinchieri, G. in *Gaseous Halos of Galaxies* (eds Bregman, J. N. & Lockman, F. J.), 215–221 (NRAO, Green Bank, 1986).
22. Scheuer, P. A. G. in *Superluminal Radio Sources* (eds Zensus, J. A. & Pearson, T. J.) 104–113 (Cambridge University Press, 1987).
23. Gopal-Krishna & Saripalli, L. *Astr. Astrophys.* **139**, L19–L21 (1984).
24. Dreher, J. W. in *Physics of Energy Transport in Extragalactic Radio Sources* (eds Bridle, A. H. & Eilek, J. A.), 109–113 (NRAO, Green Bank, 1984).
25. Saripalli, L. & Gopal-Krishna *Astr. Astrophys.* **149**, 205–208 (1985).
26. Rees, M. J. & Setti, G. *Nature* **219**, 127–131 (1966).
27. van der Laan, H. & Perola, G. C. *Astr. Astrophys.* **3**, 468–476 (1969).
28. Miley, G. K. A. *Rev. Astr. Astrophys.* **18**, 165–218 (1980).
29. Bridle, A. H. *Can. J. Phys.* **64**, 353–361 (1986).
30. Wiita, P. J. *Astrophys. J.* **221**, 436–448 (1978).
31. Smith, M. D., Smart, L., Norman, M. L. & Wilson, J. R. *Astrophys. J.* **264**, 432–445 (1983).
32. Killen, N. E. B., Bicknell, G. V. & Ekers, R. D. *Astrophys. J.* (in the press).
33. Fabbiano, G., Klein, U., Trinchieri, G. & Wielebinski, R. *Astrophys. J.* (submitted).
34. Lilly, S. J. & Prestage, R. M. *Mon. Not. R. astr. Soc.* **225**, 531–550 (1987).
35. Heckman, T. M. et al. *Astrophys. J.* **311**, 526–547 (1986).

Marine source of atmospheric acetylene

M. Kanakidou*, B. Bonsang*, J. C. Le Rouley*,
G. Lambert*, D. Martin† & G. Sennequier†

* Centre des Faibles Radioactivités, Laboratoire mixte CNRS/CEA, 91198 Gif sur Yvette, France

† Établissement d'Études et de Recherches Météorologiques, 78470 Magny les Hameaux, France

Acetylene (C_2H_2) is a non-methane hydrocarbon (NMHC) generally considered to be produced only by human activities^{1,2}. In effect, its mixing ratio is typically in the range 500–3,000 parts per 10^{12} by volume (p.p.t.v.) in inhabited countries, compared with 50–100 p.p.t.v. in remote oceanic areas. The destruction of acetylene in the atmosphere occurs only by reaction with OH radicals; acetylene is one of the longest-lived NMHC, with an average tropospheric lifetime of the order of two months, allowing this compound to reach remote areas as well as the stratosphere. Thus the presence of C_2H_2 in open oceanic atmosphere is commonly explained by its long-range transport from continental sources. Here we present and discuss new data obtained in the Indian and Pacific oceans, showing that this gas could be produced by natural processes in sea water.

The air samples were collected in evacuated stainless steel electropolished canisters. Dissolved gas in seawater samples were stripped *in situ* with ultra-grade helium by a technique described by Bonsang *et al.*³. The analysis was performed in the laboratory by gas chromatography with a flame ionization detector, by a technique adapted from Bonsang and Lambert⁴. Sample fractions were enriched in a Tenax trap at -120°C . The injection to the gas chromatograph was performed by quickly heating this trap to 140°C . The column used was a 50-m long $\text{Al}_2\text{O}_3/\text{KCl}$ porous layer open tubular capillary column (Chrom-pack). A blank was measured before each injection. Blank tests demonstrate that the helium used for stripping the sea water did not contain acetylene. A minor contamination by the glass vessel used for the extraction, ~ 5 p.p.t.v., was observed and was subtracted from the seawater concentrations. Measurements of C_2H_2 in sea water and in the atmosphere were simultaneously performed in open oceanic areas.

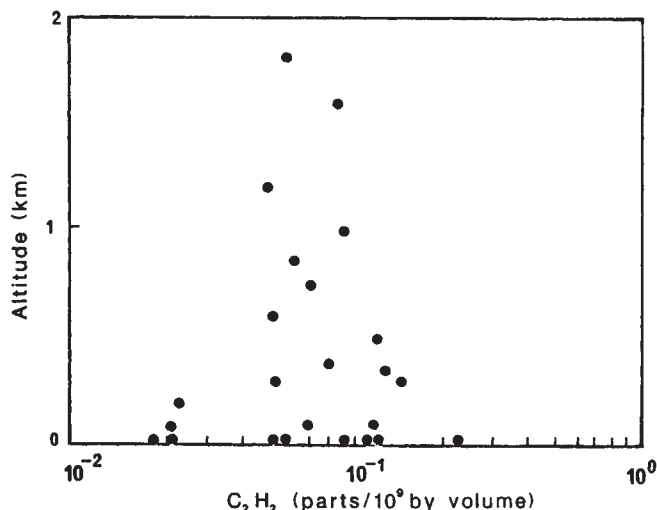


Fig. 1 Mixing ratios of acetylene in the troposphere above Hao atoll.

Table 1 reports results from three sets of air samples obtained in 1986 and 1987. The first set was sampled at Amsterdam Island (Terre Australes et Antarctiques Françaises) in the Southern Indian Ocean. The samples were collected near the background monitoring station of Pointe Benedicte, by wind blowing directly from the ocean. Simultaneously, ^{222}Rn measurements showed that we dealt with typical marine air masses.

The air samples of the second set were collected every 2 h in the Northern Central Pacific, on board RV *Researcher* during the cruise RE-87-03-RITS organized by NOAA on 28–29 May 1987. On board ship, air samples were collected to windward from the bow of the ship, through a stainless steel pipe previously rinsed with ambient air and extending beyond the hull.

A third set of air samples was obtained in June 1987 on Hao atoll (French Polynesia): sea-level samples were augmented by air samples collected at up to 1,500 m, using balloon and remote-controlled aircraft-borne devices. At the sampling place the width of the atoll was ~ 100 m and the air was blowing directly

Table 1 Mixing ratios (in p.p.t.v.) of acetylene in marine atmosphere at sea level

Site	Date	C ₂ H ₂
Amsterdam Island	12 March 1986	17
	13 April 1986	18
	12 May 1986	39
	30 May 1986	35
	13 June 1986	40
	13 July 1986	50
	12 August 1986	57
	12 September 1986	70
	15 October 1986	91
	12 November 1986	37
	23 December 1986	24
	28 February 1987	43
	27 March 1987	62
	1 May 1987	34
Hawaii	28 May 1987	111
	28 May 1987	117
	28 May 1987	116
	29 May 1987	123
	29 May 1987	107
	29 May 1987	188
	29 May 1987	201
	29 May 1987	89
	29 May 1987	77
	29 May 1987	83
Hao	29 May 1987	224
	31 May 1987	117
	31 May 1987	169
	2 June, 1987	227
	4 June 1987	56
	5 June 1987	139
	7 June 1987	125
	8 June 1987	92
	10 June 1987	110
	11 June 1987	52
	11 June 1987	22
	11 June 1987	20

from the ocean without any possible interference from local vegetation or human activities.

Table 1 shows that the mixing ratios are a little higher in the North Pacific (71–286 p.p.t.v.) than in the South Pacific (20–227 p.p.t.v.), but systematically lower in the South Indian ocean (17–91 p.p.t.v.). This observation could be in agreement with C₂H₂ production in inhabited areas of the Northern Hemisphere. But, the C₂H₂ variability actually observed in these stations, more than 300%, seems much too large to be accounted for by remote continental sources. In similar areas the variability is 0.3% for CO₂ (ref. 5), 1% for CH₄ (ref. 6), and less than 10% for CO (ref. 7) whose destruction processes by OH oxidation and lifetime are comparable to those of C₂H₂. In addition, Fig. 1 shows that the C₂H₂ variability is only $\pm 50\%$ in the free troposphere above Hao atoll as against a factor of 10 near the sea level. It therefore seems reasonable to assume the existence of a marine source of C₂H₂.

Two seawater samples were collected from the bow of

Table 2 Mixing ratios of acetylene in sea water (in 10^{-12} cm³ C₂H₂ per cm³ of sea water)

Site	Date	C ₂ H ₂ measured	Saturation concentrations	C ₂ H ₂ /C ₃ H ₈
Hawaii	28 May 1987	1,039	115	0.50
Hawaii	29 May 1987	1,282	85	0.54
Hao	4 June 1987	9,315	60	0.69
Hao	5 June 1987	2,191*	140	0.72
Hao	10 June 1987	513	110	0.37
Hao	11 June 1987	2,192	50	0.12

* Lagoon of Hao.

Researcher. Three seawater samples were obtained in open ocean ~ 3 nautical miles (~ 5.5 km) north-east of Hao atoll and another from its lagoon. Table 2 shows the results: the C₂H₂ mixing ratios in seawater vary considerably, from 513 to 9,315 p.p.t.v., but in every case the sea water was definitely supersaturated. In effect the Henry constant of C₂H₂ is about 1 in sea water at 30 °C (ref. 8). The saturation concentrations in Table 2 were established from air concentrations measured at the same place and day, shown in Table 1. It therefore seems that at least one significant (and maybe major) fraction of that C₂H₂ measured over open ocean can be ascribed to its production in sea water and its subsequent outgassing into the atmosphere. The smaller C₂H₂ concentrations measured in sub-Antarctic atmospheres could be accounted for by a smaller Henry constant in cold waters⁸.

An order of magnitude of the strength of the marine source can be tentatively evaluated from a comparison, in sea water, of the mixing ratios of C₂H₂ and C₃H₈ (propane), whose global oceanic flux was estimated by Bonsang *et al.*³ as 3 Mt of carbon per year. This gas was also measured in the same samples. Table 2 shows these results: the mixing ratios C₂H₂/C₃H₈ are between 0.1 and 0.7. Without considering the difference of the lifetimes of these gases, but taking into account the difference of composition, this figure leads to a global marine flux of C₂H₂ in the range of 0.2–1.4 Mt of carbon per year. This value is rather lower than with the C₂H₄ oceanic flux of 15 Mt of carbon per year also estimated by Bonsang *et al.*³.

Taking into account that the mean continental C₂H₂ concentration^{9,10} is two orders of magnitude higher than its marine background, we could assume that this marine source is currently weak compared with human production. But the existence of this extended marine source of C₂H₂ should be considered in that this long-lived NMHC is able to reach the stratosphere and there hinder the ozone destruction. Moreover, this marine source was possibly predominant in the past, and could account for the possible presence of C₂H₂ in air bubbles stored in ancient ices.

This study was supported by Territoire des Terres Australes et Antarctiques Françaises, Programme Interdisciplinaire de Recherches sur l'Environnement, Ministère de l'Environnement, Direction des Centres d'Expérimentations Nucléaires, and partly by US NOAA. We also thank the Greek State Organization for Scholarship for financial support (M.K.).

Received 24 December 1987; accepted 19 February 1988.

- Rudolph, J. & Ehhalt, D. H. *J. geophys. Res.* **86**, 11959–11964 (1981).
- Nelson, P. F., Quigley, S. M. & Smith, M. Y. *Atmos. Envir.* **17**, 439–449 (1982).
- Bonsang, B., Kanakidou, M., Lambert, G. *J. Atmos. Chem.* **6**, 3–20 (1988).
- Bonsang, B. & Lambert, G. *J. Atmos. Chem.* **2**, 257–271 (1985).
- Gaudry, A., Ascencio, J. M. & Lambert, G. *J. geophys. Res.* **88**, 1323–1329 (1983).

- Steele, L. P. *et al. J. Atmos. Chem.* **5**, 125–171 (1987).
- Heidt, L. E. *et al. J. geophys. Res.* **85**, 7329–7336 (1980).
- Perry, R. H. & Chilton, C. H. *Chemical Engineers' Handbook*, 5th edn (McGraw-Hill, New York, 1973).
- Greenberg, J. P. & Zimmerman, P. R. *J. geophys. Res.* **89**, 4767–4778 (1984).
- Rudolph, J., Ehhalt, D. H. & Khedim, A. *J. Atmos. Chem.* **2**, 117–124 (1984).

Superstructure of the superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ by high-resolution electron microscopy

E. A. Hewat*, M. Dupuy*, P. Bordet†, J. J. Capponi†, C. Chaillout†, J. L. Hodeau† & M. Marezio‡

* CEA, IRDI, D.LETI, CENG, 85X, 38041 Grenoble, France

† CNRS Laboratoire de Cristallographie 166X, 38042 Grenoble, France

‡ AT&T Bell, Murray Hill, New Jersey 07974, USA

The structure of the 80-K superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ (ref. 1) determined by X-ray and neutron diffraction^{2,4} is an Aurivillius-related phase as suggested by Michel *et al.*⁵, with approximate unit-cell dimensions $a = \sqrt{2}a_p$, $b = 5\sqrt{2}a_p$ and $c = 8a_p$, where the perovskite lattice parameter $a_p = 3.8 \text{ \AA}$. It consists of layers of perovskite sandwiched between BiO layers. Although electron microscopy does not have sufficient resolution to unambiguously determine the basic structural unit of such a compound, it is an invaluable tool in determining the superstructure⁶. Here we present a high-resolution electron microscope study which reveals waves of distortion along the b -axis, giving a superlattice slightly larger or smaller than $5\sqrt{2}a_p$. The superlattice is composed of building blocks four, five, and six times $\sqrt{2}a_p/2$ in approximately periodic combinations. The very low 'twin' density in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ and the nature of the 'twins' (successive planes with the a and b axes interchanged, also called twist boundaries) implies that twinning does not play an essential role in the superconducting phenomenon.

The Bi-Sr-Ca-Cu-oxides were prepared as described in Bordet *et al.*⁴, with nominal cation compositions of 2,2,2,4 and 2,2,1,4, with T_c up to 80 K. Electron microscope observations on the majority phase, prepared as described in Hewat *et al.*⁷, were performed on a JEOL 200CX at 200 kV and a JEOL 4000EX at 400 kV. These microscopes have a point resolution of 2.3 Å and 1.7 Å respectively.

$\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ is seen to be a layered compound that cleaves easily on a preferred plane. It has the slippery feel of graphite. The layers slide easily over each other. Moiré patterns and other contrast phenomenon associated with wrinkling of the layers are seen in practically all crystals.

Viewed down the c -axis, the preferred orientation of the sheets, the image is dominated by a square net of $2.7 \text{ \AA} \times 2.7 \text{ \AA}$ ($\sqrt{2}a_p/2 \times \sqrt{2}a_p/2$). In agreement with X-ray and neutron results, the unit cell is seen to be $\sqrt{2}a_p \times (\sim)5\sqrt{2}a_p$, and is centred with $cm\bar{m}$ symmetry. At certain defocus, and especially for slightly tilted crystals, it is evident that there are waves of distortion along the b -axis that give rise to the $5\sqrt{2}a_p$ periodicity. The main

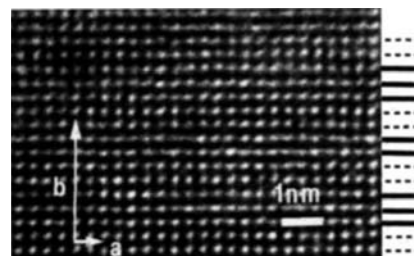


Fig. 1 [001] projection of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ crystal at a defocus which shows the sequence of dotted and continuous lines corresponding to atomic columns aligned and tilted with respect to the incident beam.

component of displacement is perpendicular to the b -axis. This is deduced from the observation that there is a repetition of two rows where the atomic columns are imaged as spots, then three rows that become continuous lines (the c -axis atomic columns are tilted about the b -axis). See Fig. 1. This is in agreement with our X-ray results which indicate that the Bi atoms are displaced along the x -axis.

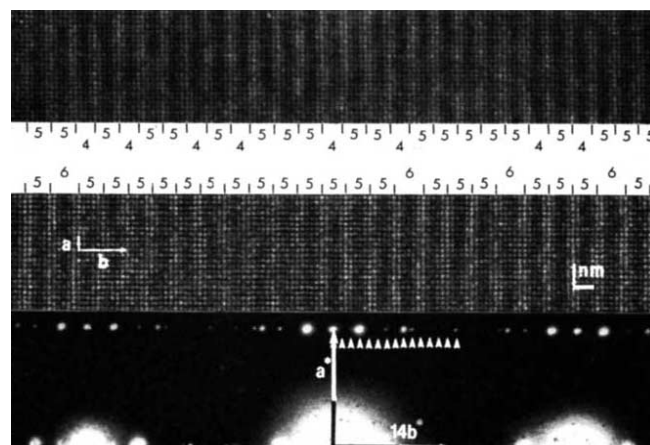
Depending on the preparation, the blocks of $\times 6$ or $\times 4\sqrt{2}a_p/2$ are inserted along the b -axis to give an average superstructure length slightly greater or less than $5\sqrt{2}a_p$. The insertion of such blocks is not perfectly periodic. This argues in favour of the existence of discrete building blocks, rather than an incommensurate wave of fixed length (Fig. 2). The diffraction patterns can often be indexed on a 14, that is $(5+5+4)$ or 19, that is $(5+5+5+4)$ times $\sqrt{2}a_p$ unit cell and so are commensurate with a large unit cell. The average repeat periods are 4.67 and 4.75 times $\sqrt{2}a_p$ respectively for these superstructures. Image simulations could help to verify our interpretation of these features of the superstructure, however the multiplicity of possible superstructures makes this very difficult at present.

We note that very thin crystals viewed down the a -axis give optical diffraction patterns in which the 150 spots appear, that is, the spots corresponding to a_p . This is consistent with the model of a one half (or a one and one half) unit cell thick layer with one layer of Bi on each surface.

Tilting about the b -axis by 10° gives the [101] projection. In this orientation the 151 reflection appears, indicating that the structure is not body-centred. The 101 reflection also appears, indicating that it is not centred on all faces. The [101] image has Pmm symmetry.

Viewed parallel to the layers, the pairs of very dark lines must correspond to layers of BiO . Bi (atomic number 83) is a very strong scatterer. It is clear that the layers cleave between the Bi

Fig. 2 a and b , [001] projection of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ crystals. The insertion of blocks of $\times 4$ and $\times 6$ ($\sqrt{2}a_p/2$) is not perfectly periodic. c , Electron diffraction pattern corresponding to a . It can be indexed on a $(5+5+4)\sqrt{2}a_p$ unit cell. Note this is not the same indexation for b as shown in all other figures.



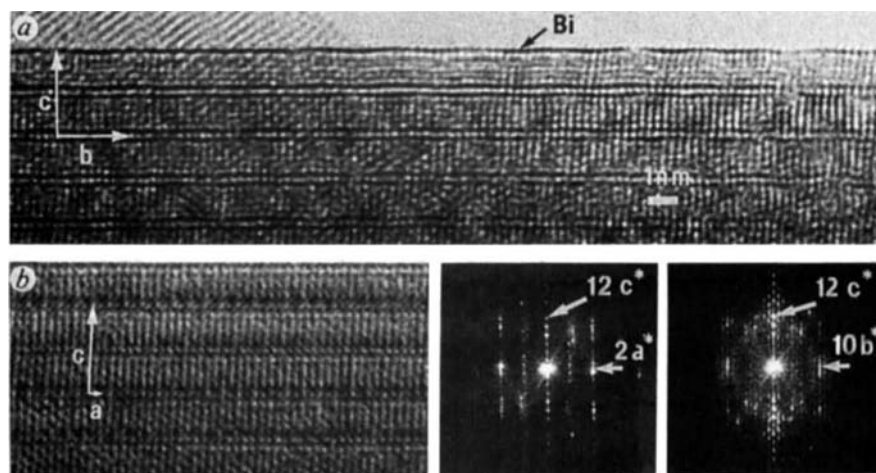


Fig. 3 *a*, Image of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ close to $[100]$. Note the undulations in the Bi layer and the single Bi plane at the surface. *b*, Image of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ close to $[010]$. *c*, Optical diffraction pattern corresponding to *b*. *d*, Optical diffraction pattern corresponding to *a*.

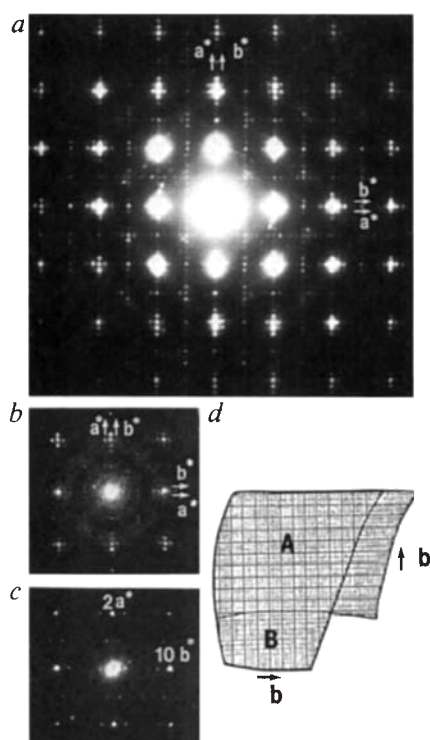


Fig. 4 *a*, Electron diffraction pattern of a twist boundary seen down $[001]$. *b*, Optical diffraction pattern from a region of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ containing a twist boundary. (This corresponds to region A of *d*.) *c*, Optical diffraction pattern from a perfect region of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ (This corresponds to region B of *d*.) *d*, Schematic representation of a twist boundary viewed down $[001]$.

bilayer, leaving a monolayer of Bi (Fig. 3*a*). The crystals seen in this orientation are usually too small to be orientated by electron diffraction. They are very often the curled-up edges of large sheets. The Bi planes are seen to undulate along the b -axis to give the superlattice. The image down the b -axis has $cm\bar{m}$ symmetry (Fig. 3*b*).

Images slightly off the short a -axis reveal fascinating moiré patterns, due to the variation of inclination of atomic columns approximately parallel to the c -axis. Depending on the exact orientation, the optical diffraction pattern is consistent with either p_{gg} or p_{gm} ($c' = c/2$) projection symmetry. It is not evident which is correct from these images.

$\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ forms non-stoichiometric stacking faults with, for example, half the usual layer thickness. It also forms twist boundaries (which have been called 'twins'), where successive

layers have the a - and b -axes interchanged. This is possible because b is approximately equal to a for the basic unit cell, see Fig. 4. The twist boundary plane is on (001) . This type of defect is often associated with overlapping crystals. The optical diffraction pattern shown in Fig. 4*b* is the same over a large region of crystal. This differs markedly from the high density of twins seen in $\text{YBa}_2\text{Cu}_3\text{O}_7$ that have (101) twin planes. The difference of twinning behaviour in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ and $\text{YBa}_2\text{Cu}_3\text{O}_7$ implies that twinning does not play an essential role in high- T_c superconductivity.

The origin of the superstructure in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ which involves waves of distortion of the BiO layers is not well understood. It could be induced by a difference in lattice parameter between the BiO layer and the perovskite layer, or by an ordering of Ca^{2+} and Sr^{2+} cations, which have atomic radii 0.94 Å and 1.10 Å respectively. But the most probable explanation is a displacement of Bi cations associated with a small number of additional oxygen atoms in the BiO layer.

With the building blocks in this system, it is possible to form many incommensurate phases that destroy the symmetry. Careful control of the preparation conditions will be essential to produce a well-ordered commensurate phase with $b = 5a$ suitable for the precise determination of the complete superstructure by X-ray and neutron diffraction.

The determination of the structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ is a good example of the complementary roles played by X-ray, neutron and electron diffraction/microscopy. X-rays can determine the position of the cations, neutrons the position of the oxygen in the basic structure, and electron microscopy the superstructure.

After completion of this work, we saw that Zandbergen *et al.*⁸ have proposed a quite different structure for the 110 K phase, which can apparently be obtained by simply annealing the present material at 885 °C for two days (J. M. Tarascon, personal communication). But Zandbergen *et al.*⁹ are in agreement with the structure presented here for the 80 K phase.

E.A.H. thanks M. Dupuy for the hospitality of his laboratory and for his personal support, and A. Bourret for making available his high-resolution electron microscope facilities.

Received 10 March; accepted 7 April 1988.

1. Maeda, H., Tanaka, Y., Fukutomi, M. & Asano, T. *Jap. J. appl. Phys.* (in the press).
2. Tarascon, J. M. *et al. Phys. Rev.* submitted.
3. Sunshine, S. A. *et al.* in preparation.
4. Bordet, P. *et al. Proc. Interlaken Conf. on High T_c Superconductors 29 February-4 March 1988* (eds Muller, A. & Olsen, P.) (in the press).
5. Michel, C. *et al. Z. Phys.* **B68**, 421 (1987).
6. Hewat, E. A. *et al. Proc. Interlaken Conf. on High T_c Superconductors 29 February-4 March 1988* (eds Muller, A. & Olsen, P.) (in the press).
7. Hewat, E. A., Dupuy, M., Bourret, A., Capponi, J. J. & Marezio, M. *Nature* **327**, 400-402 (1987).
8. Zandbergen, H. W. *et al. Nature* (submitted).
9. Zandbergen, H. W., Groen, P., Van Tendeloo, G., Van Landuyt, J. & Amelinckx, S., *Solid St. Commun.* (submitted).

Liquid crystalline suspensions of poly(tetrafluoroethylene) 'whiskers'

Thomas Folda*, Heinz Hoffmann*, Henri Chanzy* & Paul Smith*

E. I. du Pont de Nemours & Company, Inc., Central Research & Development Department, Wilmington, Delaware 19898, USA

Theoretical treatments predict that solutions of long rod-like molecules above a critical concentration should separate spontaneously into an isotropic and an anisotropic, or liquid crystalline phase¹⁻³. It has become evident that liquid crystallinity is dominated by geometrical factors³ and is not a property associated with molecules *per se*. Liquid crystalline order thus can be expected in dispersions of any kind of asymmetric objects, provided that flocculation does not occur. Well known systems displaying supramolecular liquid crystallinity are suspensions of tobacco mosaic virus⁴, haemoglobin S (ref. 5) and cellulose microcrystals⁶. Supramolecular liquid crystalline order in synthetic polymer systems has received little attention, and liquid crystallinity is commonly associated with intrinsically rigid or semirigid macromolecules or macromolecules with mesogenic side-groups^{7,8}. Here we report the formation of liquid crystalline order in suspensions of (extended chain) crystalline whiskers of the flexible macromolecule poly(tetrafluoroethylene) (PTFE).

Tetrafluoroethylene (TFE) was polymerized in a 400-ml glass reaction vessel containing 200 ml water, 10 mg potassium persulphate initiator and 2 g of the surfactant ammonium perfluorodecanoate ($C_{10}F_{19}COONH_4$). The polymerization was performed with stirring at 80 °C at a constant TFE pressure of 400 kPa for a reaction time of 5 min. The final PTFE concentration in the suspension amounted to 2.7% w/w. The polymer had an estimated molar mass of $25,000 \text{ kg kmol}^{-1}$ (determined according to ref. 9). After standing at room temperature for ~4 h, the PTFE suspension had separated into two phases. The upper phase was isotropic and showed no optical birefringence; birefringence could readily be induced in this phase by shearing the suspension, or by placing it in a magnetic field (T.F. and C. R. Fincher, unpublished observations). The polymer weight fraction in this upper phase was ~0.02. The bottom layer showed remarkably high birefringence (see optical micrograph in Fig. 1), which is indicative of the liquid crystalline order in this phase. The PTFE concentration in the bottom layer was 6% w/w.

Morphological studies indicated that polymerization of TFE under the experimental conditions described above invariably produced rod-like poly(tetrafluoroethylene) particles. A transmission electron micrograph of these supramolecular PTFE rods, isolated from the isotropic layer, is presented in Fig. 2. The rods had an almost uniform width of ~20 nm and a length that ranged from about one micrometre to several tens of micrometres. On average their axial ratio (x , length/diameter) was estimated to be of the order of 10^3 . (According to Flory's approximative relation for stable anisotropy³, a solution of rods having an axial ratio of 1,000 should exhibit an anisotropic phase at a volume fraction exceeding about 0.008. This critical value corresponds to ~1.9% by weight in the present PTFE system. The suspension described above had an overall rod concentration of 2.7% w/w, which satisfies the predicted requirement for the existence of an anisotropic phase.)

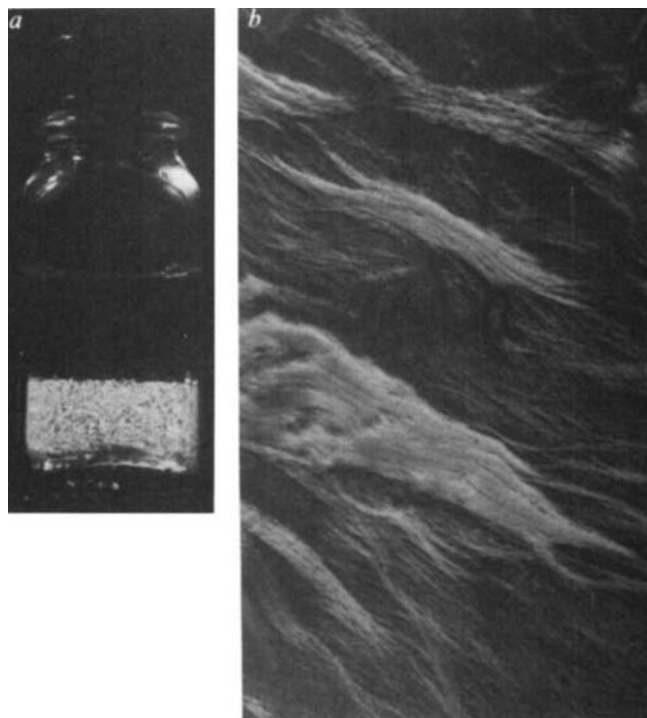


Fig. 1 Optical micrograph (crossed polarizers) of anisotropic suspension of PTFE whiskers produced by emulsion polymerization with 1% w/v of ammonium perfluorodecanoate as surfactant.

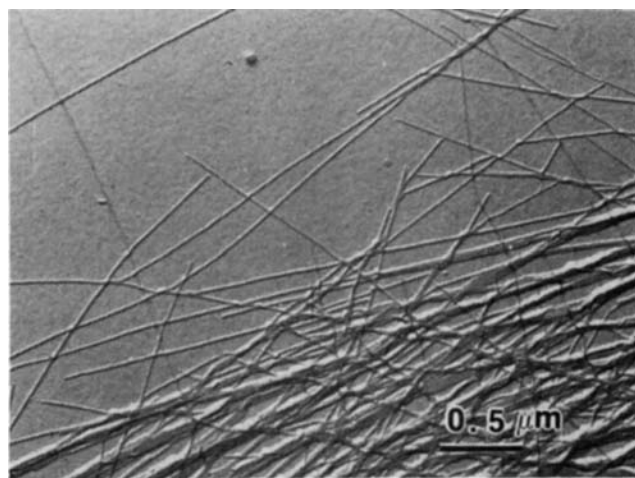


Fig. 2 Transmission electron micrograph of PTFE whiskers of Fig. 1. Specimen shadowed with W/Ta. Scale bar, 0.5 μm.

Figure 3a displays a selected area electron-diffraction pattern (insert) of the marked single rod, in the proper orientation with respect to the rod. Once indexed, this pattern indicated that the molecular axis of the polymer was parallel to that of the rod. It was previously¹⁰ inferred from electron diffraction patterns, bright-field and high-resolution images that such PTFE rods are, in fact, fully extended chain single-crystal whiskers. This is illustrated in Fig. 3b, which shows a high-resolution, (100)-lattice image of a part of a whisker (for a detailed description, see ref. 10); the uninterrupted lattice fringes illustrate the perfect parallel packing of the polymer molecules within the whisker.

Figure 4 shows a low-magnification transmission electron micrograph of the anisotropic suspension of PTFE whiskers that was dried on a microscope grid. This photomicrograph

* Present addresses: Narmco Materials, 1440 N. Kraemer Blvd., Anaheim, California 92806, USA (T.F.); Lehrstuhl für Physikalische Chemie der Universität Bayreuth, Postfach 3008, D-8580 Bayreuth, FRG (H.H.); CERMAV (CNRS), BP 68, 38402 Saint Martin d'Hères, France (H.C.); Materials Department and Department of Chemical and Nuclear Engineering, University of California, Santa Barbara, California 93106, USA (P.S.).

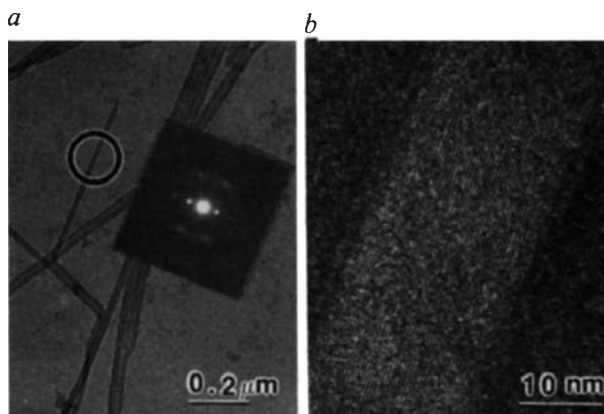


Fig. 3 *a*, As Fig. 2, but unshadowed and observed in low-dose mode. Insert: electron diffraction pattern of single whiskers. *b*, High-resolution electron micrograph, printed in reverse contrast, of a single PTFE whisker. Note the parallel lattice fringes at a spacing of 0.49 nm. The original micrograph (before photographic enlargement) was at a magnification of 46,000, an accelerating voltage of 120 kV and an accumulated electron dose of 1 electron per Å² (details in ref. 10).

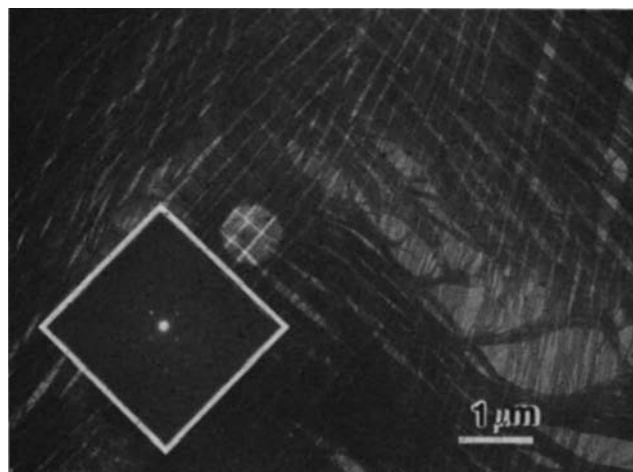


Fig. 4 Transmission electron micrograph of the dried liquid crystalline suspension of PTFE whiskers. Insert: electron diffraction of the circled area.

reveals the existence of large domains of oriented whiskers. The two domains seen in this picture are oriented roughly perpendicular with respect to one another. The insert of Fig. 4 is the electron diffractogram of the marked (bright) area. It shows a 'four point'-pattern corresponding to the two perpendicular orientations of the domains. The sharp reflections in the diffraction pattern are indicative of the local perfection of the parallel packing of the individual PTFE whiskers.

So far we have addressed the production of PTFE dispersions that showed a separation into an isotropic and an anisotropic phase. Remarkably, if the surfactant C₈F₁₇SO₃K was used in the polymerization performed under otherwise identical conditions, an entirely liquid crystalline suspension was obtained. On the other hand, if either of the two above-mentioned surfactants was used at relatively low concentrations, such as 0.1% w/w, no birefringent phase was observed; at such low surfactant concentrations, only spherical PTFE particles were formed and few, if any, whiskers were found.

Although our understanding of the production of the liquid crystalline suspensions of PTFE whiskers is not complete, it is evident (see also ref. 11) that their formation is dominated by the action of the surfactant. Currently it is thought that tetrafluoroethylene in aqueous emulsion tends to polymerize and

simultaneously crystallize into extended chain whiskers—primarily because the polymerization proceeds at temperatures far below the melting point of the final polymer¹²—provided that the growing particle is protected by surfactant against the very high polymer/water interfacial forces. The surfactant is also believed to prevent premature flocculation of the growing rods. Some indications now exist that the presence of surfactant in the polymerization, at a level above its critical concentration for rod-like micelle formation, is necessary for the production of the liquid crystalline suspensions. The influence of the chemical nature of the surfactant on the formation of entirely anisotropic suspensions or separated isotropic and anisotropic phases remains obscure for the time being.

The liquid crystalline suspensions described above illustrate once more that spontaneous ordering is not the exclusive property of solutions or melts of (semi-) rigid molecules. Here it is demonstrated that flexible synthetic macromolecules, when assembled into asymmetric objects such as the present extended-chain whiskers, may also exhibit the fascinating and useful (see ref. 13) self-alignment phenomenon that leads to liquid crystalline order. It is believed that this finding will provide new avenues for processing flexible macromolecules into highly ordered materials with anticipated superior performance.

We thank Dr C. R. Fincher for performing the magnetic field experiments and Drs A. D. English and H. W. Starkweather Jr for helpful comments.

Received 22 December 1987; accepted 23 March 1988

1. Onsager, L. *Ann. N.Y. Acad. Sci.* **51**, 627–659 (1949).
2. Ishihara, A. *J. chem. Phys.* **18**, 1446–1449 (1950); *ibid* **19**, 1142–1147 (1951).
3. Flory, P. J. *Proc. R. Soc. A* **234**, 73–89 (1956).
4. Perutz, M. F., Lignori, A. M. & Eirich, S. *Nature* **167**, 929–931 (1951).
5. Bawden, F. C. & Pirie, N. W. *Proc. R. Soc. B* **123**, 274–320 (1937).
6. Marchessault, R. H., Morehead, F. F. & Walters, N. M. *Nature* **184**, 632–633 (1959).
7. *Polymer Liquid Crystals* (eds Ciferri, A., Krigbaum, W. R. & Meyer, R. B.) (Academic, New York, 1982).
8. *Liquid Crystalline Order in Polymers* (ed. Blumstein, A.) (Academic, New York, 1978).
9. Suwa, T., Takehisa, M. & Machi, S. *J. appl. Polym. Sci.* **17**, 3253–3257 (1973).
10. Chanzy, H. D., Smith, P. & Revol, J.-F. *J. Polym. Sci. Polym. Lett.* **24**, 557–564 (1986).
11. Berry, K. L. *U.S. Patent* 2,559,750 (Du Pont, 1951).
12. Wunderlich, B. *Adv. Polym. Sci.* **5**, 568–619 (1968).
13. Schaeffgen, J. R. *et al.*, in *Ultra-high Modulus Polymers* (eds Ciferri, A. & Ward, I. M.) 173–201 (Applied Science, London, 1979).

The dispersal of the Amazon's water

Frank E. Muller-Karger*, Charles R. McClain† & Philip L. Richardson‡

* Horn Point Environmental Laboratories, University of Maryland, Box 755, Cambridge, Maryland 21613, USA

† NASA, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

‡ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

The Amazon is the largest river system in the world, contributing about 6×10^{12} m³ of fresh water to the tropical Atlantic each year^{1,2}. This is about 16% of the annual discharge into the world's oceans. Yet the fate of this water and of the dissolved and particulate material discharged with it^{3,4} has remained unclear. Previous observations of 300-km diameter lenses of Amazon water off South America^{2,5} and a plume that extends into the Atlantic^{6,7} indicate some offshore movement, but the relationship of these and along-shore currents has remained obscure. New information, obtained with NASA's Coastal Zone Color Scanner (CZCS) and with drifting buoys, reveals that the discharge of the Amazon is carried offshore around a retroflection of the North Brazil Current and into the North Equatorial Countercurrent towards Africa between June and January each year. From about February to May the countercurrent and the retroflection weaken or vanish, and Amazon water flows northwestward towards the Caribbean Sea.

We studied the dispersal of Amazon water with 32 CZCS satellite images collected between November 1978 and December 1982, and with trajectories of surface drifters released from ships between 1983 and 1986. Visible radiance backscattered out of the upper optical depth of the ocean was estimated from the CZCS radiance⁸⁻¹⁰. Pigment concentrations (chlorophyll *a* plus phaeopigments and dissolved organic material) were estimated from the water-leaving radiance by using empirical relationships based on ratios of the blue or blue-green (443 or 520 nm) relative to the green (550 nm)⁸⁻¹⁰ radiances. Previous studies have shown that in relatively clear oceanic water, pigment estimates are within 30–40% of average near-surface phytoplankton pigment concentrations^{9,11}. Nearshore, pigment concentrations are less reliable and were considered only qualitatively. Inshore of the 10 m isobath directly off the Amazon delta, high pigment concentrations ($>5 \text{ mg m}^{-3}$; Figure 1) reflect turbidity due to suspended sediment⁴. A permanent phytoplankton bloom occurs between the 10 and 30 m isobaths (pigment up to 30 mg m^{-3} ; P. Dustan personal communication), where suspended sediments in surface waters decrease to $<10 \text{ mg}$ sediment per liter⁴, allowing light to penetrate deeper. Currents off the delta advect this water and an abundant flora of neritic diatoms¹² northwestward along the coast of Brazil.

A comparison of the CZCS images and tracks of surface-drifting objects¹³ north of the Amazon delta provided a consistent picture of a fairly regular seasonal cycle. The path followed by the Amazon water during February–May was different from that traced during June–January (Fig. 1). During February–May, Amazon water flowed into the Caribbean in a broad (150–200 km), continuous band of high concentrations of pigment ($>1.5 \text{ mg m}^{-3}$) and sediment (Fig. 1, *Top*). Offshore, a large area in the tropical Atlantic contained pigment concentrations $>0.2 \text{ mg m}^{-3}$, compared with the $<0.15 \text{ mg m}^{-3}$ background concentrations in oceanic water. This is a result of earlier offshore movement of shelf water during June–January. During February–May, surface drifters in the tropical Atlantic moved westward until they were within a few hundred km of the coast, where they turned and drifted more northwestward. The trajectories implied a continuous boundary current with characteristic speeds of $\sim 90 \text{ cm s}^{-1}$.

During June–January, five buoys veered offshore near 5°N 50°W and continued eastward, following a meandering pattern (Fig. 1, *Bottom*). This retroflexion of the North Brazil Current (NBC) is the western source of the North Equatorial Counter-current (NECC). A similar pattern was observed in the June–January CZCS images, indicating that Amazon water was being carried offshore around the retroflexion and into the NECC. Pigment concentrations in the plume were highest near the continental margin ($>1.5 \text{ mg m}^{-3}$) and decreased to about 0.4 mg m^{-3} 1,000 km into the NECC. The patterns seen during June–January were similar to the distribution of sea-surface salinity (Fig. 2), and agreed with historical ship drifts¹⁴ and numerical models¹⁵. The continuity in the distribution of pigments showed the Amazonian origin of the offshore plume and is evidence that the whole near-surface NBC retroflects.

The shape of the retroflexion seems to be due to an overshoot by the NBC past the mean latitude of the NECC (8°N), followed by a quasistationary wave train downstream (Figs 1 and 3). The images revealed that the northern crest of the retroflexion frequently reached 11°N , 53°W , and that a southern meander trough formed near 4.5°N , 46°W . A second oscillation was evident with a trough near 4.5°N , 39°W . The average wavelength of the meanders was 730 km. The centre of the retroflexion was near 8°N , 51°W , where a depression in the thermocline has been observed repeatedly¹⁶.

From August to October, buoys in the NBC drifted past the Amazon mouth at speeds of about 90 cm s^{-1} and then retroflected, reaching 45°W in about 18 days. From here they gradually slowed to 30 cm s^{-1} as they drifted eastward. About half the buoys in the NECC continued eastward into the Guinea Current

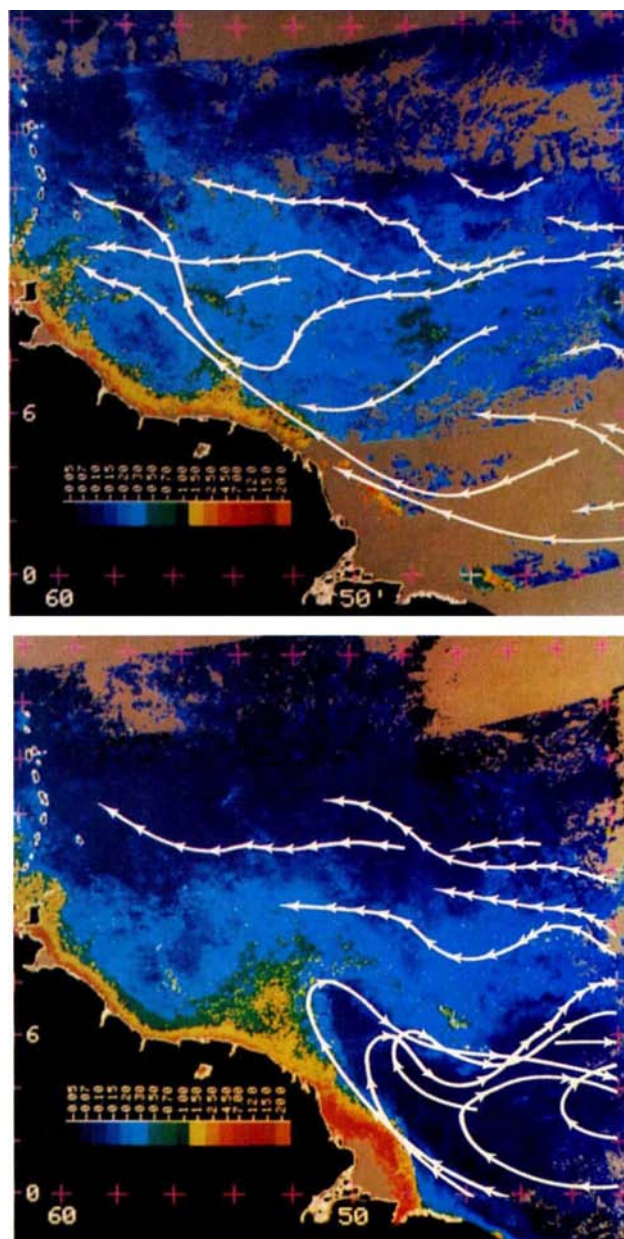


Fig. 1 Seasonal composites of CZCS images and surface drifters for the western tropical Atlantic. The CZCS composites show the mearpigment value [mg m^{-3}] of cloud-free pixels in sets of images collected during 1979–80. Land was masked black and clouds or missing data grey. Smoothed trajectories of buoys tracked from 1983 to 1986 (ref. 13) were added to the composites. Arrowheads are spaced at five-day intervals. Some trajectories were omitted to reduce clutter. *Top*, a composite of five images from 15 January to 20 February 1980 and trajectories for January–June. The Amazon plume appears as a wide yellow-orange band along the coast. Offshore, the blue-green colours indicate the region over which Amazon water from the previous season is dispersed. *Bottom*, a composite of 15 images from 21 July 1979 to 9 January 1980 and trajectories for July–December. The plume is carried offshore near $6\text{--}8^\circ \text{N}$ as seen by yellow and green values fading to light blue in contrast to dark blue oceanic water.

near Africa. The other half moved northward into the North Equatorial Current (NEC) and then westward.

These circulation patterns indicate that Amazon water discharged during June–January (about 60% of the annual discharge) may take between a few months and a year to reach

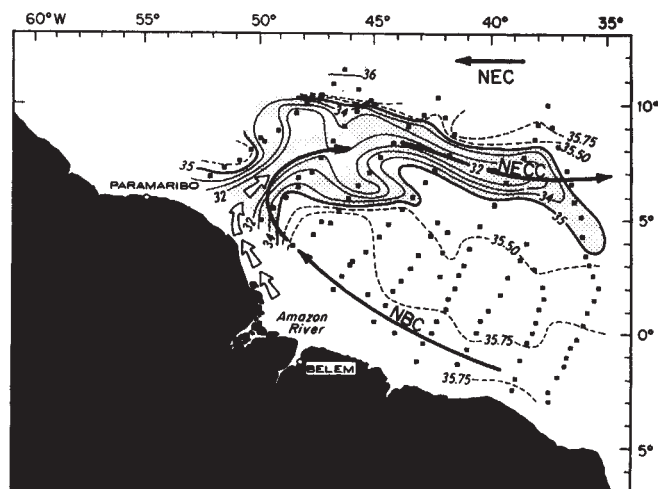


Fig. 2 Quasi-synoptic map of sea-surface salinity for the period 24 July to 11 September 1964 (ref. 7). Arrows were added to show the inferred path of low-salinity Amazon water and major currents.

the Caribbean. Also, while the discharge of one year drifts westward in the NEC, water entrained in a newly formed NECC may mix with some of the older discharge. A fraction of Amazon water may continue moving eastward and reach Africa, as indicated in a global composite of CZCS imagery for December 1981 (G. C. Feldman, personal communication).

The patterns of plume dispersal vary from year to year (Fig. 3). For example, during February–May of 1979, the retroflexion elongated to the northeast as seen in a CZCS composite of May 1979¹⁷. Also, during June–January of 1980 and 1981, some Amazon water was lost to the northwest, away from that retroflexing into the NECC (Fig. 3). Spatial variability is so large in this region that coarse sampling from ships could lead to misconceptions about the circulation.

This information is conclusive evidence that the low salinity (34.8–35.0‰) observed across the tropical Atlantic between 5° and 10° N (ref. 18) is not solely due to precipitation associated with the Intertropical Convergence Zone (ITCZ)¹⁹. But the processes that lead to the formation of the retroflexion of the NBC and the offshore transport of Amazon water are not clear. Numerical models indicate that the NECC forms owing to variation in the curl of the wind stress associated with the ITCZ and geostrophic adjustment of the water^{15,20}. In addition, the NBC, which has negligible planetary vorticity near the Equator, flows into a region of waters with positive vorticity²¹. This combination of factors may cause the pronounced anticyclonic turn of the NBC. Similar processes may be responsible for the seasonal eastward turn of the Somali Current²².

The patterns described above help explain the variability of salinity and pigments observed at Barbados²³. Also, the greater flux of organic particulates to the bottom of the tropical Atlantic relative to the tropical central Pacific²⁴ may be explained if the Amazon plume contains abundant phytoplankton. But it is still not clear how such large amounts of phytoplankton could remain viable in the plume. There are several likely sources of nutrients for phytoplankton. Upwelling takes place inshore of the retroflexion^{5,25,26} and also possibly along crests in the NECC, where geostrophic adjustment occurs. Recycling of nutrients within the plume probably plays a major role, and grazing by zooplankton, sinking, and mixing with nutrient-poor oceanic water limit the horizontal extent of the blooms. None of these variables has been studied yet.

The many different paths taken by Amazon water depend on the variability of the retroflexion and of the NECC. Owing to

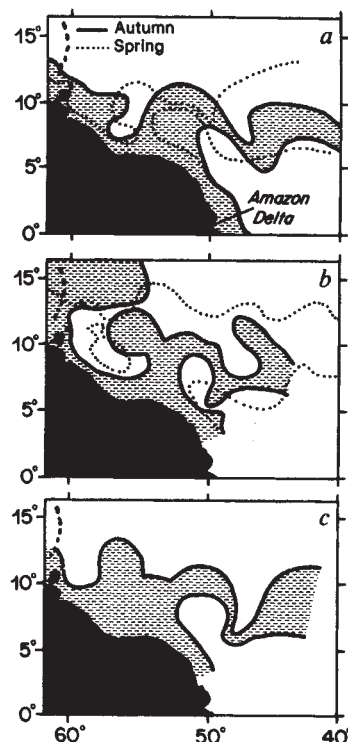


Fig. 3 Schematic representation of the Amazon plume as observed in seasonal composites of CZCS images for (a) 1979, b, 1980 and c, 1981. No images were available for February–May 1981. Pigment contours follow approximately the 0.3 mg m^{-3} isopleth and were smoothed for clarity. Stippled areas represent the June–January (northern autumn) extent of the plume, broken lines show February–May conditions (northern spring).

the large scale of the processes observed, future studies in the western tropical Atlantic need to be multidisciplinary and done from platforms that provide synoptic information. Such studies need to address plankton dynamics to arrive at a better understanding of the geochemical effects of the Amazon River on the interior Atlantic Ocean.

We thank John Sissala, Mike Doline and Richard Sipes (Nimbus Operations, General Electric) for their help in selecting the imagery; Judy Chen, Mike Darzi and Jim Firestone (General Sciences Corporation) for implementing the image-processing system (SEAPAK); and Tom Fisher and Bill Schmitz for helpful comments on the manuscript. This work was supported by the NASA Graduate Student Researcher's Program, the Ocean Processes Branch of NASA Headquarters, and the National Science Foundation.

Received 16 November 1987; accepted 29 February 1988.

1. Sioli, H. in *River Ecology* (ed. Wretton, B. A.) 461–488 (Blackwell, London, 1975).
2. Gibbs, R. J. *J. mar. Res.* **28**, 113–132 (1970).
3. Milliman, J. D. & Meade, R. H. *J. Geol.* **91**, 1–21 (1983).
4. Nittrouer, C. A., Curtin, T. B. & DeMaster, D. J. *Cont. Shelf Res.* **6**, 151–174 (1986).
5. Ryther, J. H., Menzel, D. W. & Corwin, N. J. *J. mar. Res.* **25**, 69–83 (1967).
6. Perloth, I. in *Proc. Symp. Oceanogr. Fisheries Resources trop. Atlantic 185–191* (UNESCO, Paris, 1966).
7. Cochran, J. D. *J. mar. Res.* **27**, 327–334 (1969).
8. Clark, D. K. in *Oceanography from Space* (ed. Gower, J. F. R.) 227–237 (Plenum, New York, 1981).
9. Gordon, H. R. *et al. Appl. Opt.* **22**, 20–36 (1983).
10. Gordon, H. R. *et al. Appl. Opt.* **22**, 3929–3931 (1983).
11. Gordon, H. R. *et al. J. mar. Res.* **40**, 491–502 (1982).
12. Wood, E. J. *Bull. mar. Sci.* **16**, 102–123 (1966).
13. Richardson, P. L. & Reverdin, G. *J. geophys. Res.* **92**, 3691–3708 (1987).
14. Richardson, P. L. & Walsh, D. J. *geophys. Res.* **91**, 10537–10550 (1986).
15. Philander, S. G. H. & Pacanowski, R. C. *J. geophys. Res.* **91**, 14192–14206 (1986).
16. Bruce, J. G., Kerling, J. L. & Beatty, W. H. III *Prog. Oceanogr.* **14**, 57–63 (1985).
17. Esaias, W. E., Feldman, G. C., McClain, C. R. & Elrod, J. A. *Eos* **67**, 835–837 (1986).

18. Henin, C., Hisard, P. & Piton, B. *FOCAL* Vol 1 (ORSTOM, Paris, 1986).
19. Neumann, G. *Deep Sea Res. Suppl.* 16, 165-177 (1969).
20. Busalacchi, A. J. & Picaut, J. *J. phys. Oceanogr.* 13, 1564-1588 (1983).
21. Csanady, G. T. *J. mar. Res.* 43, 553-579 (1985).
22. Bruce, J. G. *J. phys. Oceanogr.* 14, 826-832 (1984).
23. Borstad, G. A. *J. mar. Res.* 40, 421-434; 435-452 (1982).
24. Pliskaln, C. H. & Honjo, S. *Global biogeochem. Cycles* 1, 31-48 (1987).
25. Mazeika, P. A. *Di. hydrogr.* Z. 26, 49-73 (1973).
26. Gibbs, R. J. *Deep Sea Res.* A27, 857-866 (1980).

D/H ratios of environmental water recorded by D/H ratios of plant lipids

Leonel da Silveira Lobo Sternberg

Department of Biology, University of Miami, Coral Gables, Florida 33124, USA

Isotope ratios of chemical components are powerful tools in the interpretation of palaeoenvironments, particularly carbonates from foraminiferans^{1,2}, desert caliche³, and desert pavements⁴. Isotope ratios of plant cellulose are also indicators of environmental variables such as temperature and relative humidity⁵⁻⁷. Several workers have reported climatic fluctuations based on hydrogen and oxygen isotope ratios of cellulose from tree trunks⁸⁻¹⁰ and peats^{11,12}. Here I present measurements that demonstrate that for submerged aquatic plants δD values of lipids record D/H ratios of environmental water, whereas cellulose does not. I further demonstrate that δD values of lipids in conjunction with $\delta^{18}O$ values of cellulose provide significant information on isotope ratios of environmental water.

Oxygen and hydrogen isotope ratios of plant cellulose were first proposed as palaeoclimatic indicators because isotope ratios of rainwater (meteoric water as defined by Craig¹³) are related to climate; that is, rainwater from warmer regions at lower latitudes or altitudes has high D/H and $^{18}O/^{16}O$ ratios, whereas in cooler regions at higher latitudes or altitudes these ratios are lower. Plants record the hydrogen and oxygen isotope ratios of water available for growth in their cellulose⁵⁻⁷. Thus deducing climate through isotope ratios of meteoric water is a matter of deciphering the isotopic fractionations that take place during water uptake, evapotranspiration and incorporation into cellulose. The major problems with the use of cellulose to interpret climate, however, are the effect of evapotranspiration in modifying isotopic ratios of leaf water relative to ground water^{7,14}, and species-specific variable isotopic fractionations that occur during metabolic processes responsible for incorporation of hydrogen in cellulose¹⁵⁻²⁰. Particularly relevant to this paper is the latter problem. Large variability in δD values of non-exchangeable hydrogens from cellulose of different species of plants grown in a single site, exposed to the same climate and meteoric water, have been observed^{15,16}. This variability also extends to aquatic plants, indicating that variable hydrogen isotopic fractionations during cellulose synthesis are primarily due to fractionations occurring during biochemical reactions^{17,18}. In addition, fractionations occurring during biochemical reactions are influenced by environmental factors such as temperature¹⁹ and light quality²⁰. As a result of this variability, absolute δD values of water available during cellulose synthesis cannot be deduced from hydrogen isotope ratios of cellulose. Differences in hydrogen isotope ratios of plant cellulose through time can be due to changes in δD values of groundwater, to differences in isotopic fractionations particular to each species, or to environmental effects on the hydrogen isotope fractionations occurring during cellulose synthesis. Thus climatic information from hydrogen isotope ratios of cellulose cannot be precisely estimated.

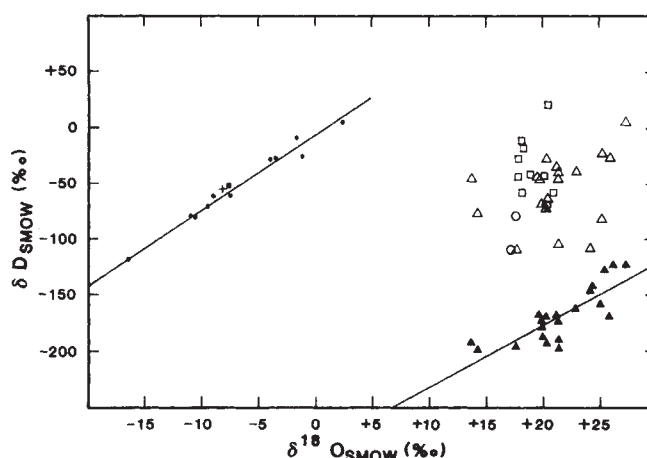


Fig. 1 The relationship between δD and $\delta^{18}O$ values of lake waters, δD values of non-exchangeable hydrogen of cellulose and $\delta^{18}O$ values of cellulose from submerged aquatic plants, and δD values of lipids and $\delta^{18}O$ values of cellulose. Symbols: ●, isotope ratios of lake water in this study; +, ■, isotope ratios of waters measured by DeNiro and Epstein¹⁹ and Epstein *et al.*⁵ respectively; Δ, isotope ratios of cellulose measured in this study; □, ○, isotope ratios of cellulose measured by DeNiro and Epstein¹⁹ and Epstein *et al.*⁵ respectively; ▲, δD values of lipids versus $\delta^{18}O$ values of cellulose. Correlation coefficients, r , were 0.95 for water values, 0.38 for cellulose values, and 0.80 for lipids versus cellulose values.

There are three possible solutions to this problem. The first and most difficult is a complete understanding of all isotopic fractionations occurring during water uptake through cellulose synthesis for different species. This knowledge can then be used to deduce δD values in meteoric water from isotope ratios of plant cellulose. The second is to find a single species whose isotopic fractionations do not vary between individuals of the same species or with environmental conditions. The third solution, specifically addressed here, is the analysis of a plant component that incorporates deuterium without the metabolic effects observed for cellulose. Previous D/H ratio measurements in lipids and cellulose from plants show that whereas cellulose has a large variation in δD relative to the δD of the available water, lipids have a relatively constant δD value^{21,22}. Thus lipids may be a plant metabolic component that truly records the δD values of environmental waters.

The feasibility of using δD values of lipids in conjunction with $\delta^{18}O$ values of cellulose to deduce isotope ratios of meteoric water is demonstrated here. Submerged aquatic plants from several lakes of different geographical regions having water with different isotope ratios were collected and $^{18}O/^{16}O$ and/or D/H ratios of their cellulose and lipids, and the surrounding water, were determined (Table 1). Submerged aquatic plants are ideal organisms for this analysis because they do not transpire, and transpiration modifies the isotope ratio of water available for cellulose synthesis relative to the water source¹⁴. δD and $\delta^{18}O$ values of lake waters were highly correlated with each other ($r = 0.95$, $P < 0.01$, Fig. 1) and this relation is expressed by the regression equation

$$\delta D_{\text{water}} = 6.7\delta^{18}O_{\text{water}} - 6.9 \quad (1)$$

The slope of this regression line (6.7) is smaller than that of meteoric water (8.0) (ref. 13). As lakes of warmer regions at lower altitudes evaporate more rapidly, they may have a disproportionate increase in $\delta^{18}O$ values of the water relative to increases in δD values due to kinetic effects, thus decreasing the slope of the line. δD values of extracted and nitrated plant cellulose were not related to the δD values of lake water

Table 1 δD and $\delta^{18}O$ values of lake waters at different geographical locations and altitudes, δD and $\delta^{18}O$ values of cellulose and δD values of lipids from plants growing in the respective lakes

Species	Locations	Water		Cellulose		Lipid δD
		δD	$\delta^{18}O$	δD	$\delta^{18}O$	
<i>Myriophyllum</i> sp.	Villeperdue, France, 100 m	-25	-1.2	-107	+24.2	-146
<i>Alisma plantago</i>	Lespuau, France, 300 m	-27	-3.5	-81	+25.0	-158
<i>Potamogeton crispus</i>	Lespuau, France, 300 m	-27	-3.5	-39	+22.8	-162
<i>Myriophyllum</i> sp.	Lago Calzado, Peru, 4,425 m	-119	-16.4	-77	+14.3	-199
<i>Eleocharis acicularis</i>	Lago Calzado, Peru, 4,425 m	-119	-16.4	-46	+13.7	-192
<i>Eleocharis acicularis</i>	Imbabura, Ecuador, 3,650 m	-62	-7.4	-104	+21.3	-173
<i>Elatina</i> sp.	Imbabura, Ecuador, 3,650 m	-62	-7.4	-46	+19.9	-178
<i>Potamogeton</i> sp.	Imbabura, Ecuador, 3,650 m	-62	-7.4	-41	+21.4	-197
<i>Eleocharis maculosa</i>	Cotapaxi, Ecuador, 4,100 m	-62	-8.9	-63	+20.4	-170
<i>Equisetum boyotensis</i>	Cotapaxi, Ecuador, 4,100 m	-62	-8.9	-68	+19.8	-174
<i>Myriophyllum quitensis</i>	Cotapaxi, Ecuador, 4,100 m	-62	-8.9	-45	+19.6	-168
<i>Potamogeton panamensis</i>	Cotapaxi, Ecuador, 4,100 m	-62	-8.9	-71	+20.2	-189
<i>Calitriche nubigena</i>	Junin, Peru, 4,425 m	-71	-9.5	-108	+17.7	-196
<i>Isoetes palmerii</i>	Junin, Peru, 4,425 m	-80	-11.0	-27	+20.4	-193
<i>Crassula pallidosa</i>	Lago Chisaca, Colombia, 3,620 m	-79	-10.6	-35	+21.2	-168
<i>Hydrilla verticillata</i>	Lago Gatun, Panama, 165 m	-28	-3.9	-45	+21.4	-190
<i>Utricularia</i> sp.	Florida, USA, ≤ 1 m	+5	+2.4	—	+26.2	-123
<i>Eleocharis acicularis</i>	Florida, USA, ≤ 1 m	-8	-1.6	+6	+27.2	-122
<i>Cabomba caroliniana</i>	Florida, USA, ≤ 1 m	-8	-1.6	0	+24.3	-141
<i>Hydrilla verticillata</i>	Florida, USA, ≤ 1 m	-8	-1.6	-26	+25.8	-170
<i>Najas guadalupensis</i>	Florida, USA, ≤ 1 m	-8	-1.6	-21	+25.3	-127

Isotope ratios of water, cellulose and lipids were determined as previously described^{18,19,22} and reported in δ units, where δ (‰) = $[(R_{\text{sample}}/R_{\text{standard}}) - 1] \times 1,000$, and R is ratio of rare to common isotope. The standard used here for δD and $\delta^{18}O$ values is standard mean ocean water (SMOW).

($r = 0.36$, $P > 0.05$). This result contradicts the reported constant 22‰ difference between δD values of extracted and nitrated cellulose from terrestrial plants and the δD values of water available for plant growth⁶. δD values of lipids and $\delta^{18}O$ values of cellulose were highly correlated with the respective isotope ratios of lake waters ($r = 0.79$, $P < 0.01$ for lipids; $r = 0.94$, $P < 0.01$ for cellulose). However, interpretation of these relations is difficult because the δD and $\delta^{18}O$ values of the water when cellulose and lipids were synthesized could have been different from those at the time of water sampling. A way of overcoming this problem is to determine whether δD values of lipids or cellulose correlate with $\delta^{18}O$ values of cellulose and whether the relation between δD and $\delta^{18}O$ values of these plant metabolites is similar to the relation observed between δD and $\delta^{18}O$ values of water available for growth.

δD values of non-exchangeable hydrogen of cellulose and $\delta^{18}O$ values of cellulose for several submerged aquatic plants measured here and reported elsewhere^{5,19} are not correlated ($r = 0.38$, $P > 0.05$, Fig. 1). In contrast, δD values of lipids are correlated with $\delta^{18}O$ values of cellulose at a highly significant level ($r = 0.80$, $P < 0.01$, Fig. 1) with the linear regression equation

$$\delta D_{\text{lipid}} = 5.5 \delta^{18}O_{\text{cell}} - 286\% \quad (2)$$

The slope of this regression line (5.5) is less than that of lake water (6.7). The relation between δD of lipids and $\delta^{18}O$ of cellulose of submerged aquatic plants, including a lower slope, can be explained by deriving a theoretical equation relating the δD values of lipids to $\delta^{18}O$ values of cellulose. Consider the two equations showing the isotopic labelling of lipids and cellulose by deuterium and oxygen-18 respectively from the water available to the plant,

$$\delta D_{\text{lipid}} = \alpha_d \delta D_{\text{water}} + (\alpha_d - 1) \times 1,000 \quad (3)$$

and

$$\delta^{18}O_{\text{cell}} = \alpha_o \delta^{18}O_{\text{water}} + (\alpha_o - 1) \times 1,000 \quad (4)$$

where α_d and α_o are defined as $(D/H)_{\text{lipid}}/(D/H)_{\text{water}}$ and $(^{18}O/^{16}O)_{\text{cellulose}}/(^{18}O/^{16}O)_{\text{water}}$ respectively. Although these

terms are usually reserved for equilibrium fractionation processes, I use them here in the same sense as that used by Epstein and colleagues^{5,7}. Their use is in part justified by recent evidence that oxygen isotope ratios of cellulose may be determined by fractionations occurring during the equilibrium between hydrated carbonyl groups in carbohydrate intermediates and water during cellulose synthesis²³. Merging equations (1), (3) and (4) the following equation expressing the relation between δD of lipids and the $\delta^{18}O$ values of cellulose is derived:

$$\delta D_{\text{lipid}} = (6.7(\alpha_d/\alpha_o)\delta^{18}O_{\text{cell}}) + [1,000(\alpha_d - 1) - (6.9\alpha_d - 6,700(\alpha_d/\alpha_o)(\alpha_o - 1))] \quad (5)$$

having a slope of $6.7\alpha_d/\alpha_o$ and intercept at $[1,000(\alpha_d - 1) - 6.9\alpha_d - 6,700(\alpha_d/\alpha_o)(\alpha_o - 1)]$. Thus the slope of a plot of δD values of lipids against $\delta^{18}O$ values of cellulose is not only a function of the slope of the line described by plotting δD against $\delta^{18}O$ values of water, but of the fractionation constants α_d and α_o . Accordingly, the greater the discrimination against deuterium during the formation of plant lipid or the smaller the discrimination against oxygen-18 during cellulose synthesis, the lower the slope of this line should be. Using the α_d of 0.870 as reported for saponifiable lipids of green algae²¹, and the α_o of 1.027 as previously reported for cellulose^{5,19} in equation (5), the following is derived:

$$\delta D_{\text{lipid}} = 5.7 \delta^{18}O_{\text{cell}} - 289\% \quad (6)$$

nearly identical to the regression equation (2) obtained for the observed values.

The determination of absolute δD values of environmental water by analysis of plant lipids will allow a more precise estimation of palaeotemperatures, because it can be used to confirm the estimation of oxygen isotope ratios of water from oxygen isotope ratios of cellulose. Furthermore, with the method described here, it will be possible to determine the relation between δD and $\delta^{18}O$ values of water from lakes, which may be an important indicator of relative humidity. Because the lipid fraction of plants tends to be preserved longer it may be possible to determine climate further back in time than with cellulose.

This will be possible once it has been shown that isotopic ratio alterations due to diagenetic effects are minimal. If the results reported here can be extrapolated to terrestrial plants, then the technique of using δD values of lipids along with $\delta^{18}O$ values of cellulose may be applicable in determining the D/H and $^{18}O/^{16}O$ ratios of environmental water from δD and $\delta^{18}O$ values of lipids and cellulose from peats, where there is a large variation in isotopic ratios due to species composition^{11,12}.

I thank Peter Swart for technical support and use of his mass spectrometer, Martine Hossaert, Sharon Herman and Jon Keeley for providing samples of submerged aquatic plants, and David Lee for editorial comments. This research was supported by the Petroleum Research Fund.

Received 6 November 1987; accepted 24 February 1988.

1. Boyle, E. A. *Geochim. cosmochim. Acta* **50**, 215–276 (1986).
2. Emiliani, C. *Science* **154**, 851–856 (1966).

3. Cerling, T. E. *Earth planet. Sci. Lett.* **71**, 229–240 (1984).
4. Dorn, R. I. & DeNiro, M. J. *Science* **227**, 1472–1474 (1985).
5. Epstein, S., Thompson, P. & Yapp, C. J. *Science* **198**, 1209–1215 (1977).
6. Epstein, S., Yapp, C. J. & Hall, J. *Earth planet. Sci. Lett.* **30**, 241–251 (1976).
7. Yapp, C. & Epstein, S. *Geochim. cosmochim. Acta* **46**, 955–965 (1982).
8. Burk, R. L. & Stuiver, M. *Science* **211**, 1417–1419 (1981).
9. Krishnamurthy, R. V. S. & Epstein, S. *Nature* **317**, 160–162 (1985).
10. Stuiver, M. & Braziunas, T. F. *Nature* **328**, 58–60 (1987).
11. Schiegl, W. E. *Science* **175**, 512–513 (1972).
12. Brenninkmeijer, C. A. M., Geel, B. van & Mook, W. G. *Earth planet. Sci. Lett.* **61**, 283–290 (1982).
13. Craig, H. *Science* **133**, 1702–1703 (1961).
14. Dongman, G., Nurnberg, H., Forstel, H. & Wagener, K. *Radiat. Envir. Biophys.* **11**, 41–52 (1974).
15. Sternberg, L. & DeNiro, M. J. *Science* **220**, 947–948 (1983).
16. Sternberg, L., DeNiro, M. J. & Johnson, H. B. *Pl. Physiol.* **74**, 557 (1984).
17. Sternberg, L., DeNiro, M. J. & Keeley, J. E. *Pl. Physiol.* **76**, 69–70 (1984).
18. Keeley, J. E., Sternberg, L. O. & DeNiro, M. J. *Aquat. Bot.* **26**, 213–223 (1986).
19. DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **45**, 1885–1894 (1981).
20. Estep, M. & Hoering, T. C. *Geochim. cosmochim. Acta* **44**, 1197–1206 (1980).
21. Sternberg, L. da S. L., DeNiro, M. J. & Ajie, H. O. *Planta* **169**, 320–324 (1986).
22. Sternberg, L., DeNiro, M. J. & Ajie, H. *Phytochem.* **23**, 2475–2477 (1984).
23. Sternberg, L. da S. L., DeNiro, M. J. & Savidge, R. A. *Pl. Physiol.* **82**, 428–431 (1986).

Glacial tunnel valleys and Quaternary history of the outer Scotian shelf

Ron Boyd, D. B. Scott & M. Douma

Centre for Marine Geology, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

High-resolution seismic reflection data indicate the presence of huge sub-surface channel networks on the outer Nova Scotian continental shelf. Channel axes extend to over 450 m below present sea level (b.s.l.). Channel walls average 2–3 km in width. Mechanisms capable of producing such channels include fluvial, submarine canyon or glacial erosion. Considerable debate has focused on the positions of the Tertiary/Quaternary (T/Q) and Pleistocene/Holocene (P/H) boundaries^{1–4} in this region and the relationship of the T/Q boundary to the unconformity at the base of the channel networks. Uncertainty also surrounds the extent of Pleistocene ice sheets on the south-east Canadian continental margin^{4–7}. Here we present a new stratigraphy for the Sable Island region (Fig. 1) based on seismic profiles, with lithologic and biostratigraphic control provided by two strategically placed boreholes. This stratigraphic analysis establishes the positions of the T/Q and P/H boundaries outside the valleys at 51 m and 220 m b.s.l. respectively. Our analysis also implies that the large channel systems are tunnel valleys, cut by a sub-ice meltwater process under a pre- to early Wisconsinan ice sheet which extended close to the shelf edge.

Seismic reflection profiles were acquired with minisleeve exploder and 12-kJ sparker sources recording twelvefold coverage down to 2 s of two-way acoustic travel time. Wave-equation migration techniques were applied to lines near the borehole giving a maximum vertical resolution of 2 m. Seismic stratigraphic methods⁸ were used to define seismic sequences and facies. Seismic lines spaced 400 m apart were run parallel to the northern shoreline of Sable Island, with the closest line ranging 1–3 km from the island and 2 km from the C67 and SAB 85 boreholes (Fig. 2). The SAB 85 borehole was drilled on Sable Island along the axis of a major channel observed on marine seismic profiles north of the island and correlated under the island using exploration land seismic profiles. The C67 borehole was drilled outside the channel, 3 km east of SAB 85. Coring and continuous sampling were conducted to a total depth of 152 m in SAB 85. The C67 borehole was a standard oil exploration well with samples recovered every 3 m through the Cenozoic section. Cores from SAB 85 were split on site, described, photographed and tested for geotechnical properties⁹. Borehole samples from C67 were mainly used for foraminiferal

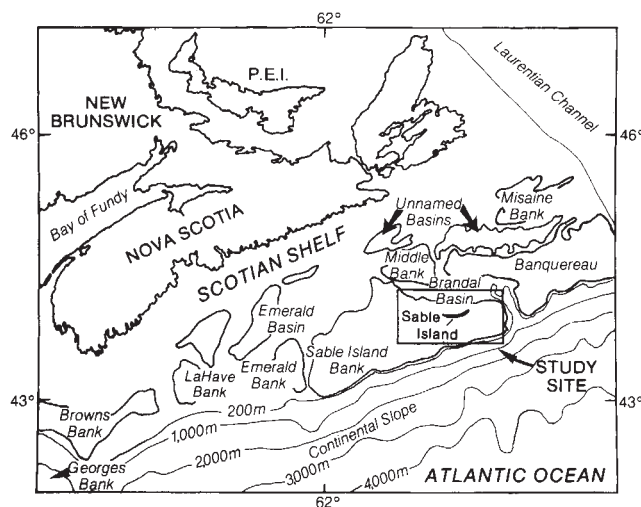


Fig. 1 The Scotian shelf is located on the south-east Canadian continental margin. Topography is dominated by a series of mid-shelf basins and outer shelf banks, one of which is Sable Island Bank, the site of a sub-surface glacial tunnel valley network and several recent boreholes.

analysis, sediment grain size and composition. Gamma logging was performed on both boreholes and electric logs were also available from C67. Identification of sedimentary units employed a standard facies logging method supplemented for SAB 85 by sieve grain size analyses and X-radiographs. Benthic foraminifera were used as palaeoenvironmental indicators^{10,11}.

The stratigraphy of the two boreholes and the area around Sable Island was derived from a combination of biostratigraphy, lithostratigraphy, seismic stratigraphy and radiocarbon-dating techniques (Table 1). Micropalaeontological analysis of C67 samples indicates that the T/Q boundary lies at a depth of 220 m (Fig. 3). This boundary was picked on the basis of the first downgoing occurrence (K. MacKinnon and F. Gradstein personal communication) of the Tertiary marker species *Astergerina guruchi*¹². The position of the boundary coincides with a decrease in the number of reworked pre-Quaternary microfauna and lies within the uppermost extreme of several coarsening upward sequences inferred from gamma logs and lithology. These sequences are within seismic facies 1 (Table 1) which is characterized by parallel-to-divergent and occasionally shingled clinoforms. The shingled clinoforms and associated small channels indicate deltaic conditions. Additional support for such

conditions is provided by the presence of terrestrial plant fragments and lignite within the sequences. Together these features suggest continued deposition across the T/Q boundary in a prograding shelf setting. An increase in reworked Cretaceous/Tertiary (K/T) microfauna above the boundary signals the onset of Pleistocene glaciations. The T/Q boundary lies within a conformably bedded sequence dipping 1–2° to the south-east, thus indicating the presence of a wedge of early Pleistocene sediment thickening south from Sable Island to the shelf edge.

Radiocarbon dating on clays at 28.5 m in SAB 85 yielded a date of $3,365 \pm 70$ yr BP. Shells from 72 m yielded a date of $37,210 \pm 400$ yr BP. Earlier studies¹³ have shown that few reworked K/T microfaunas occur in Holocene sediments. Studies of Pleistocene core material from near this area¹⁴ have shown that the amount of reworked microfauna in glacial marine silts has a positive correlation with glacial erosion. Abundant reworked K/T microfaunas in muds below 56 m in SAB 85 indicate the presence of glacial conditions and hence a Pleistocene age. These data together with seismic interpretations and earlier Sable Island ¹⁴C dates¹⁵ place the H/P boundary at or above 51 m in SAB 85; this boundary is also the base of a major coarsening upward sand sequence (Fig. 3) referred to on the Scotian Shelf as the Sable Island sand and gravel¹⁶. The sediments below 51 m in SAB 85 and between 51 m (upper boundary picked from log and seismic correlation—see Fig. 3) and 220 m in C67 are therefore of Pleistocene age.

Analysis of marine seismic reflection data collected close to Sable Island (Fig. 4) shows the channel penetrated by the SAB 85 borehole to be asymmetric in cross-section with a maximum channel axis depth of 432 m below sea level (all depth conversions assume an average sediment velocity of $1,600 \text{ m s}^{-1}$). The steeper western channel wall has an average slope of 30–35°. The narrow central section of the channel occupied by seismic sequence 2A is only 1.8 km wide while the broader channel fill

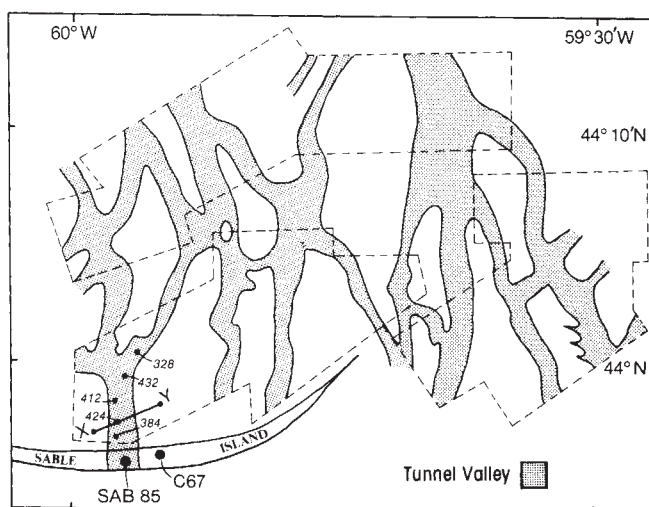


Fig. 2 The glacial tunnel-valley drainage pattern near Sable Island forms a complex network unlike that resulting from fluvial processes. These data are derived from site surveys (boundaries in thin dashed line) using multichannel seismic methods and extrapolated with single channel regional seismic reflection profiles. On Sable Island exploration land seismic data was used to locate the tunnel valley. Also shown are the borehole locations of Fig. 3, the seismic profile (X-Y) of Figure 4 and maximum channel depths (metres below sea level) in the tunnel valley near SAB 85.

of sequence 2B is 3.8 km wide (Fig. 4). The channel axis has a thalweg of variable depth which shows no consistent trend of deepening in any direction, a maximum along-channel relief of 40 m over 200 m (1:5) and tributary valleys with base levels 105 m shallower than the main channel. Seismic facies within the channel fill are chaotic in the lower part (sequence 2A on

Table 1 Stratigraphy of the Sable Island area

Seismic stratigraphy	Lithostratigraphy	occasional bioturbation, cross lamination and erosional truncation. Coarsening-up cycles	Biostratigraphy
Sequence 1—composed of several lower order sequences separated by highly continuous, high amplitude unconformities. Low-mod. amplitude, high continuity reflectors with parallel to divergent configuration and shingled clinoforms. Reflector R_3	50–60 m thick between R_2 and seafloor. Shallow depth and muting inhibit internal features. C67	77–64 m and 56–0 m. Mature quartz sand, usually <10% shell, granules, organic matter and mud.	SAB 85
Sequence 2A—linear, asymmetric channel fill geometry, 260 m thick, bounded by high relief unconformities. Internal reflectors have low continuity, variable amplitude, chaotic configuration. Seismic pullup below. Reflector R_{3B}	Facies 1: 0–98 m, clear-white mature quartz sand. Low mud content, occasional shells (<i>Ostrea</i> sp.) Gamma response 2 coarsening up cycles, Electric(E) log alternating increase/decrease upward.	Facies 2: 77–104 m and 143–152 m, interbedded sand silt and clay. Upper boundary where mud consistently >10%, other boundaries at massive clay of facies 3. Coarsening-up trend 104–77 m. Scoured bases and coarsening and fining up cycles. Minor components shell and mud clasts, soft sediment deformation. Undrained shear strengths of 90–250 kPa.	28.5 m <i>Elphidium</i> sp. and planktonics. ¹⁴ C date $3,365 \pm 70$ yr BP
Sequence 2B—channel fill geometry, 110 m thick. Broader than sequence 2, occurs as cap over channel axis. Basal boundary is erosional unconformity, flat upper boundary between 50–60 m.b.s.l. Upper sequence has channel cut and fill, lower part is opaque. Seismic pulldown below. Reflector R_2	Facies 2: 98–177 m, fine-to-coarse-grey sand with red clay stringers, minor shells, plant fragments and muddy matrix. E log and gamma highly variable.	Facies 3: 104–143 m and 64.5 m, stiff olive-grey silty clay. Structureless with little bioturbation. Minor wispy laminations, sand and silt patches, sediment deformation and erosional scours. Normally consolidated to slightly pre-consolidated. Undrained shear strengths 180–300 kPa, water content 20–40%, liquid limit 45, plastic limit 21.	56.3 m Reworked K/T fauna 57.8 m <i>Elphidium</i> – <i>Cassidulina</i> assemblage with K/T forms.
Sequences 3 and 4—generally	Lithostratigraphy SAB 85		64.5 m Reworked K/T fauna. 67.7 m <i>Elphidium excavatum</i> fauna. 71.9 m ¹⁴ C date 37,210 ± 400. 80.4 m <i>Elphidium</i> fauna with K/T forms. 90–137 m <i>Elphidium</i> and <i>Cassidulina</i> fauna and high K/T%.
	Facies 1: 0–77 m, fawn grey sand, thin mud, gravel and organic horizons. Structureless with		137–140 m Barren. 141–152 m Low K/T%.
			C67 0–200 m High % of K/T reworked forms, low abundance. 200 m High abundance of Miocene/Pliocene fauna. T/Q boundary at 220 m based on first occurrence of <i>Astergerina guruchi</i> .

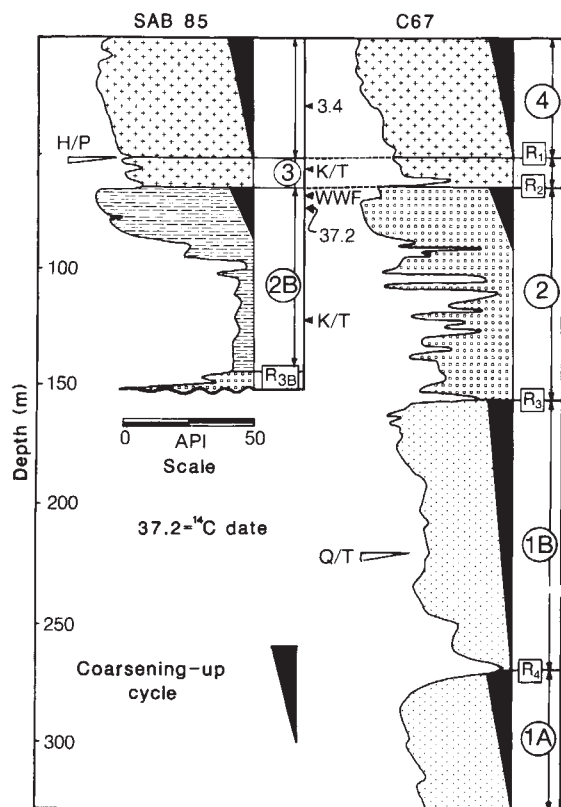


Fig. 3 Gamma log traces from the SAB 85 and C67 boreholes (location in Fig. 2) are primarily a result of natural radioactivity emitted from fine-grained clay minerals and are analogous to vertical profiles of grain size. The C67 borehole shows two coarsening upward cycles below 150 m interpreted as deltaic in origin, one of which contains the Quaternary/Tertiary (Q/T) boundary. The Holocene/Pleistocene (H/P) boundary lies at or above 51 m, the base of an upward coarsening sand body in the SAB 85 borehole. Symbols within circles refer to seismic sequences, symbols within squares are seismic reflector sequence boundaries (see Fig. 4 and Table 1). Radiocarbon dates from SAB 85 and surrounding boreholes (quoted in kyr BP) provide a Holocene/late Wisconsinan chronostratigraphy. K/T indicates the presence of reworked Cretaceous/Tertiary microfauna and hence a glacial period. WWF indicates the presence of a warm-water microfauna and hence an interglacial period.

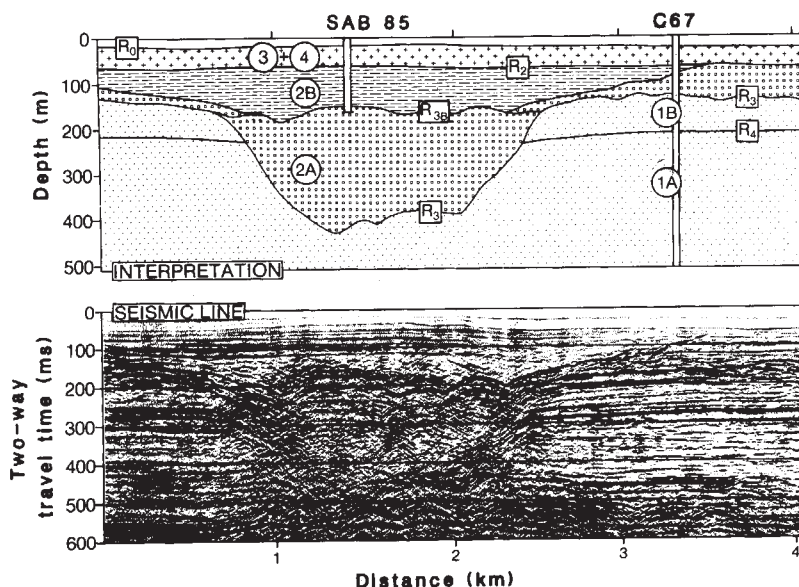
Fig. 3) but are opaque to reflector-free in the overlying channel fill of sequence 2B that has been shown from foraminiferal and grain size data to be glacial marine clay (Table 1).

The characteristics of the channel penetrated by SAB 85 are shared by several other major sub-surface channels in the region. Channel-axis depths are commonly >300 m but range to maximum depths of 450 m. The R3 unconformity outside the major channels generally has small-scale relief of <100 m and channel tops lie between 80 and 190 m b.s.l. All channels are relatively narrow, have steep valley walls, variable depth along channel, hanging tributary valleys and chaotic basal fill seismic facies. The channels form an unusual interconnected network (Fig. 2) in the Sable Island region. Within this network, bifurcation and rejoining are common with major channels becoming smaller and then subsequently expanding along the channel axis.

Such channel characteristics are unlike submarine canyons, which occur on the continental slope or close to the shelf break. Submarine canyons consistently deepen downslope and have a tributary pattern but no bifurcation on the upper continental slope. The Sable Island region is located on the continental shelf over 30 km from the Pleistocene or late Tertiary shelf edge. The channel depths and other characteristics also do not fit a fluvial origin. The T/Q boundary lies below the R₃ channel base unconformity (Table 1), establishing the channels as Pleistocene in age and not Tertiary as previously suggested¹. Oxygen isotope and sea-level data¹⁷ predict that Pleistocene low sea level stands never extended deeper than approximately 200 m. This precludes subaerial rivers of Pleistocene age from incising channels with axial depths of over 400 m below present sea level. The channel network patterns and along-channel gradients are unlike any which commonly result from fluvial drainage.

However, the channel networks are remarkably similar to those created by sub-ice tunnel valleys^{18,19}. This interpretation also fits well with the established Pleistocene age of the channels, their great depth and along-channel gradients. The most comprehensive previous description of Pleistocene tunnel valleys comes from northern Germany¹⁸. There, channels are incised up to 434 m below sea level, have widths up to 4 km, steep side walls with slopes >55°, and display rises and depressions along the valley length as well as interconnected channel networks. The German tunnel valleys are infilled at the base with chaotic coarse-grained sediments including slump blocks. The upper valley fills are fine-grained clays similar to those of lithofacies 3 (Table 1) in SAB 85 and are interpreted as glacial lacustrine deposits. The Sable Island channels are therefore also interpreted as subglacial tunnel valleys, similar to features found in

Fig. 4 Multi-channel seismic profile from 2 km north of Sable Island (location in Fig. 2). Interpretation shows seismic sequences (circles), sequence boundaries (squares) and projected position of SAB 85 and C67 boreholes. The principal feature seen is a 424-m-deep tunnel valley below the SAB 85 borehole. The lower boundary of the tunnel valley is defined by the R₃ reflector and infilled by sequence 2A.



northern Germany, and on a smaller scale in Minnesota¹⁹ and the southern North Sea²⁰. The microfauna of lithofacies 2 and 3 indicate that at least the upper part of the tunnel valley fill was deposited under marine conditions. The presence of a warm-water *Elphidium* fauna overlying the K/T dominated ice margin fauna in facies 2, the 37,210 kyr date at 72 m and the dating of the last major glacial period on the inner to mid Scotian shelf as oxygen isotope stage 4 (ref. 21) constrains the youngest age for tunnel valley incision to either the early Wisconsinan (~60–75 kyr) or to an even earlier Pleistocene glaciation. Other estimates suggest that the Illinoian glaciation was the most extensive on the Eastern Canadian margin²².

Theoretical considerations of subglacial tunnel discharge support the concept that channels cut into the substrate occur under active ice²³. Larger channels are predicted to grow at the expense of smaller ones²⁴. Sudden discharge events cause a rapid rise in subglacial water pressure and the development of brief periods of unstable sheet flow followed by channel cutting²⁵. These theoretical predictions are supported by the large scale of the Sable Island tunnel valleys. The presence of an erosional unconformity (R_3) cut on interfluvies between channels (Fig. 4) suggests the initial occurrence of a high-volume regional discharge which localized in the channels during a short-lived event.

The scale of the Sable Island channels is sufficiently large as to suggest that this event was a catastrophic process. Assuming even a relatively low current velocity of 1 ms^{-1} , the bankfull discharge (to the level of R_{3B}) through the Sable Island tunnel valley beneath SAB 85 would have been $6.4 \times 10^5 \text{ m}^3 \text{ s}^{-1}$ (or $2 \times 10^4 \text{ km}^3 \text{ yr}^{-1}$). Placed in perspective, this figure is a factor of three greater than the annual discharge of the Amazon River; and such a rate of discharge could empty a large glacial lake the size of Lake Ontario in less than one month. Ice sag into the channel roof is not considered in these calculations and would act to decrease discharge. If all seven of the tunnel valleys of similar dimensions in the vicinity of Sable Island (Fig. 2) discharged concurrently or flow velocities exceeded 1 m s^{-1} , the volume would be correspondingly higher. Such huge discharge volumes could not be sustained for long periods, and even a short-lived discharge event would require sub-ice storage of a substantial meltwater volume. The presence of such large meltwater volumes has been predicted theoretically²⁴ and confirmed experimentally beneath the Antarctic ice cap by radio echo-sounding techniques²⁶. Catastrophic release of meltwater is common on a smaller scale subglacially as jökulhlaups such as those found beneath Icelandic glaciers^{27,28} and on a larger scale from glacially dammed ice contact lakes (for example, the Lake Missoula or Snake River Pleistocene flood events²⁹). The Sable Island channel bases are cut far enough below sea level to require a driving mechanism other than gravity-forced drainage down slope. The most likely mechanism appears to be pressure from the overlying ice acting to squeeze the meltwater out between the ice sheet and the substrate. The factor which actually triggers the discharge event remains to be determined, but may be related to sealevel rise. Storage of large meltwater volumes requires a suitable subglacial depression to act as a reservoir. Potential reservoirs are present on the Scotian shelf in mid-shelf basins (Fig. 1) such as Brandal Basin, Emerald Basin and an unnamed series of small basins north of Banquereau which contain water depths of up to 316 m. The volume of the combined basins to the north of Sable Island which could act as reservoirs is $\sim 490 \text{ km}^3$, sufficient to maintain 1 m s^{-1} bankfull discharge through the single tunnel valley beneath SAB 85 for nine days.

Mobil Oil Canada kindly provided the processed seismic data and logistical support necessary for drilling the SAB 85 borehole. Shell Canada Resources provided additional seismic data. Jacques-McClelland Geosciences performed the drilling and geotechnical testing for SAB 85. Funding for the project was provided by Canadian government NSERC grants to R.B. and D.B.S. and by the Atlantic Geoscience Centre of the Canadian

Geological Survey. K. MacKinnon and F. Gradstein identified the C67 microfauna. The manuscript benefited from critical reviews by R. Dalrymple, G. Fader and J. Shaw.

Received 4 September 1987; accepted 25 January 1988.

- King, L. H., MacLean, B. & Fader, G. B. *Can. J. Earth Sci.* **11**, 89–100 (1974).
- Hardy, I. A. *Geol. Surv. Can. Pap.* **74-30**, 163–274 (1974).
- Amos, C. L. & Knoll, R. G. *Am. Ass. Petrol. Geol. Bull.* (in the press).
- Piper, D. J. W. et al. in *Geological Survey of Canada, Geology of Canada* (eds Keen, M. J. & Williams, G. L.) Vol. 2, Ch. 11 (Geological Survey of Canada, Ottawa, in the press).
- Flint, R. F. *Glacial and Quaternary Geology* (Wiley, Toronto, 1971).
- Grant, D. R. *Geographie Physique et Quaternaire* **31**, 247–260 (1977).
- Prest, V. K. *Geol. Surv. Can. Pap.* **84-10**, 21–36 (1984).
- Vail, P. R. et al. *Am. Ass. Petrol. Geol. Mem.* **26**, 49–212 (1977).
- Jacques-McClelland Geosciences Tech. Rep. G042, Halifax 1985 (Centre for Marine Geology, Dalhousie University, Halifax, 1985).
- Scott, D. B., Schafer, C. T. & Medioli, F. S. *J. foram. Res.* **10**, 205–234 (1980).
- Scott, D. B., Mudie, P. J., Vilks, G. & Younger, C. D. *Mar. Micropaleontol.* **9**, 181–218 (1984).
- Gradstein, F. M. & Agterberg, F. B. in *Quantitative Stratigraphic Correlation* (eds Cubitt, J. M. & Reymont, R. A.) 119–170 (Wiley, New York, 1982).
- Williamson, M. A., Keene, C. E. & Mudie, P. J. *Mar. Micropaleontol.* **9**, 219–239 (1984).
- Scott, D. B. & Medioli, F. S. *Abstracts with Program, 23rd NE Geological Society of America, Annual Meeting*, Portland, **68** (1988).
- Scott, D. B., Medioli, F. S. & Duffett, T. E. *Geology* **12**, 173–176 (1984).
- King, L. H. *Mar. Sci. Pap.* **1** (Department of Energy, Mines and Resources, Canada, 1970).
- Beard, J. H., Sangree, J. B. & Smith, L. A. *Am. Ass. Petrol. Geol. Bull.* **66**, 158–169 (1982).
- Ehlers, J. *Ann. Geol.* **2**, 143–146 (1981).
- Wright, H. E. Jr *Geol. Soc. Am. Mem.* **136**, 251–276 (1973).
- Dingle, R. V. *Geol. Mijnbouw* **50**, 679–686 (1971).
- King, L. H. & Fader, G. B. *Geol. Surv. Can. Bull.* **363**, 72 pp (1986).
- Alam, M. & Piper, D. J. W. *Geographie Physique et Quaternaire* **31**, 15–22 (1977).
- Shaw, J. *Society of Economic Paleontologists and Mineralogists Short Course* **16** (eds Ashley, G. M., Shaw, J. & Smith, N. D.) 7–84 (SEPM, Tulsa, 1985).
- Shreve, R. L. *J. Glaciol.* **11**, 205–214 (1972).
- Rothlisberger, H. *J. Glaciol.* **11**, 177–203 (1972).
- Oswald, G. K. A. & Robin, G. de Q. *Nature* **245**, 251–254 (1973).
- Thorarinsson, S. *Geogr. Annaler* **21**, 216–242 (1939).
- Clague, J. J. & Mathews, W. H. *J. Glaciol.* **12**, 501–504 (1973).
- Bretz, J. H. *J. Geol.* **77**, 505–543 (1969).

Gold and native copper in supergene sulphides from the Mid-Atlantic Ridge

M. D. Hannington*, G. Thompson†, P. A. Rona‡ & S. D. Scott*

* Department of Geology, University of Toronto, Toronto, Canada M5S 1A1

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ NOAA/AOML, 4301 Rickenbacker Causeway, Miami, Florida 33149, USA

In 1985 black smokers and massive sulphide deposits were discovered in the TAG hydrothermal field in the rift valley of the Mid-Atlantic Ridge near latitude $26^{\circ}08' \text{ N}$ (ref. 1). Sulphide samples from chimneys and mounds were recovered by dredging at depths between 3,620 and 3,670 m. Mineralogical and chemical analyses of the samples reveal (1) primary, unaltered Zn-Fe-Cu-sulphides containing 0.8–4.0 p.p.m. Au (0.02–0.12 oz per ton) and (2) secondary Cu-rich sulphides with native copper, formed by supergene alteration of primary Cu-Fe-sulphides, containing up to 16.4 p.p.m. Au (0.48 oz per ton) as free native gold. This is the first reported occurrence of secondary enrichment of gold and copper in recent submarine sulphides. The high gold grades and native copper associated with secondary Cu-sulphides resemble occurrences in some supergene gossans overlying ancient massive sulphide deposits on land².

The sulphides in the TAG hydrothermal field occur in a mound $\sim 220 \text{ m}$ by 250 m across and up to 50 m in height. A temperature of 305° C for discharging hydrothermal fluid was measured with a temperature probe in one active chimney at the summit of the mound during operations with the research submersible *Alvin* in 1986 (ref. 3). Primary sulphides in samples

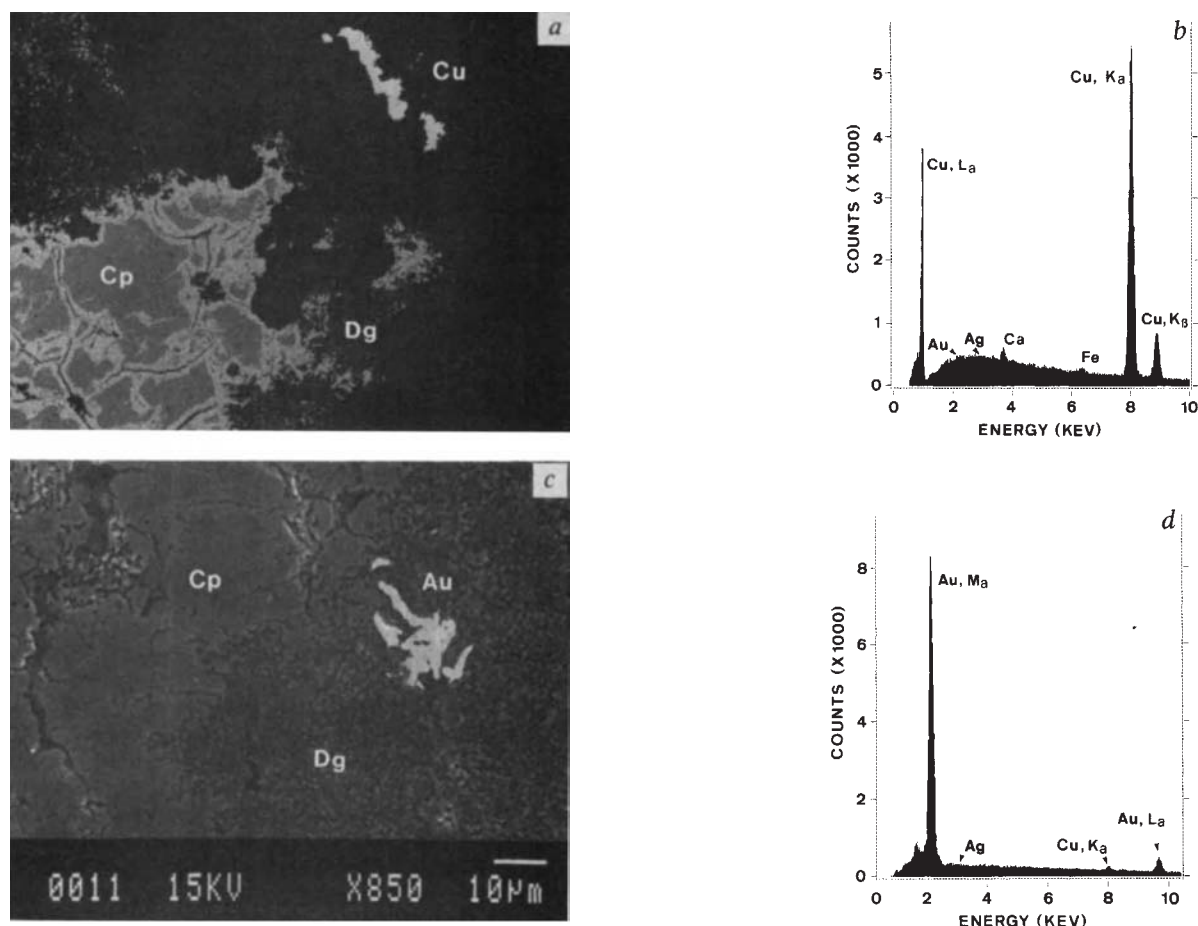


Fig. 1 *a*, Native copper (Cu) in amorphous silica and carbonate (black) near altered chalcopyrite (Cp) and secondary digenite (Dg) (back-scattered electron image). *b*, Energy dispersive X-ray spectrum of native copper. Ca and Fe X-rays are from the surrounding carbonate and Fe-oxides. The copper contains no detectable Au or Ag. *c*, Native gold (Au) in disseminated digenite (Dg) surrounding altered chalcopyrite (Cp) (secondary electron image). *d*, Energy dispersive X-ray spectrum of native gold, showing trace Cu. The gold contains no detectable Ag.

recovered from the mound include marcasite, pyrite, sphalerite, chalcopyrite and minor bornite, together with amorphous silica and aragonite^{1,3-6}. Secondary sulphides consist of Cu-rich minerals: digenite, covellite and bornite after primary chalcopyrite. These are associated with corroded pyrite, abundant late-stage amorphous silica, Fe-oxides and minor atacamite ($\text{Cu}_2\text{Cl}(\text{OH})_3$) (ref. 4). Digenite is the principal secondary Cu-sulphide, replacing chalcopyrite first along fractures within grains and subsequently around the margins of Cu-Fe-sulphides. Fe-oxides coat samples of primary and secondary sulphides but, as there are no pseudomorphous textures after primary sulphide grains, they do not appear to have been derived from the oxidation of pre-existing sulphides.

Table 1 lists eight representative chemical analyses from 19 analyses performed on different types of samples by a combination of neutron activation and X-ray fluorescence. Primary Zn-Fe-Cu-sulphides contain an average of 2.4 p.p.m. Au ($n=13$) with a median value of 3.0 p.p.m. Au, whereas secondary Cu-sulphides contain an average of 11.2 p.p.m. Au ($n=6$), with a median value of 11.4 p.p.m. Au. No free gold was observed in the primary sulphides at $\times 1,250$ magnification of polished specimens, suggesting that gold is present only as grains smaller than about 1 μm . Free grains of native gold up to 15 μm across occur with fine-grained digenite at the margins of altered chalcopyrite in the secondary Cu-sulphides (Fig. 1). A total of four grains of gold were identified all with the same textural relationship to the secondary Cu-sulphides in two samples selected for petro-

graphic work. The gold here is nearly pure, containing less than 10 weight percent Cu, and no detectable Ag. This is in contrast to gold in primary massive sulphides which usually contains some Ag. Native copper occurs as free, irregular grains up to 50 μm across within late-stage amorphous silica near the boundary between altered Cu-sulphides and the Fe-oxides that coat the samples (Fig. 1). Native copper is closely associated with crystals of atacamite in most of the samples. The native copper is much more common than native gold and is locally coarse enough to be visible without magnification.

Primary Cu-Fe-sulphides (for example, 1-44) are relatively Au-poor, with less than 1 p.p.m. Au, but are enriched in Mo and Co compared to other primary sulphides. Au-rich, primary Zn- and Fe-sulphides contain up to 3.6 p.p.m. Au, 95 p.p.m. Ag and 272 p.p.m. As, and resemble gold-rich sulphides found at Axial Seamount and Southern Explorer Ridge in the north-eastern Pacific Ocean⁷. Secondary Cu-sulphides contain up to 16.4 p.p.m. Au and 152 p.p.m. Ag with higher Au/Ag ratios than in the primary sulphides. Fe-oxides (such as 1-18) are anomalously enriched in As, but contain virtually no Au or Cu.

Gold and copper are released from primary sulphides during oxidation by ambient seawater or dilute hydrothermal fluid and are locally enriched by chemical migration and redeposition within secondary sulphides at the margins of altered primary Cu-Fe-sulphide grains. This process is suggested by the absence of coarse-grained gold in the Cu-Fe-sulphides, which indicates that dissolution and reprecipitation of gold must have occurred

Table 1 Gold, silver and selected trace metal contents of representative samples

Number	Primary Fe sulphides		Primary Cu-Fe-sulphides	Primary Zn-Fe sulphides		Secondary Cu-sulphides		Fe oxides 1-18
	1-10	1-16	1-44	1-8	1-45	1-26	1-46	
Au p.p.m.	1.62	4.00	0.83	3.62	3.03	5.65	16.4	0.03
Ag p.p.m.	17	38	14	95	37	45	116	<10
As	50	121	28	272	49	44	64	156
Sb	4	10	5	30	14	14	24	14
Mo	75	51	197	67	45	21	286	33
Co	2	3	32	3	3	3	3	4
Cd	50	90	20	490	200	40	10	<10
Se	<10	<10	13	<10	<10	<10	<10	<10
Cu wt%	0.09	2.74	25.0	0.17	0.16	32.5	39.0	0.10
Fe	41.6	25.1	29.0	29.3	18.8	18.3	20.2	50.7
Zn	2.09	3.25	0.52	17.4	7.20	1.19	0.63	0.33
Pb	0.10	0.05	<0.01	0.13	0.06	0.02	0.02	0.05
S	52.1	29.2	35.8	40.4	29.1	26.4	22.6	0.18

Au, Ag, As, Sb, Mo, Co, Cd, Se, Fe, Zn by neutron activation analysis (M.D.H.); Cu, Pb, S by X-ray fluorescence (B. Schroeder, Woods Hole Oceanographic Institution⁶).

after deposition of the primary sulphides. In addition, the absence of detectable silver in secondary gold grains from the altered sulphides is in contrast to the silver-gold alloys typically found in primary hydrothermal sulphides^{2,11,12}. The complexity of speciation, where drastically contrasting chemical environments (that is, reduced primary sulphides and oxygenated seawater) are intimately juxtaposed in time and space, makes predictions about the behaviour of gold and copper in supergene fluids difficult. But at the low pH produced by oxidation of chalcopyrite, gold may be dissolved from the primary sulphides as auric chloride complexes (AuCl_4^-) (refs 8-11). Reprecipitation of the gold may occur as a result of increasing pH or reduction to the native metal by ferrous iron¹¹. The different stabilities of gold, silver and copper chlorides in solution can account for the precipitation of high-purity gold and the effective separation of native gold and copper^{11,12}. The coprecipitation of native copper and Cu chlorides (namely atacamite) attests to the importance of chloride complexing for metal transport and the low pH of the oxidizing solutions¹⁰. No *in situ* precipitation of Fe-oxides occurs at the site of sulphide oxidation, and the loss of indigenous iron from the oxidized chalcopyrite could be responsible for the Fe-oxide coating on most of the samples.

The association of secondary Cu-sulphides, free gold, and native copper at the TAG hydrothermal field is currently unique among modern sea floor sulphide deposits. The well developed secondary processes in the TAG deposit are in keeping with other observations that the TAG hydrothermal field is mature. The large TAG deposit has grown through a combination of coalescing chimneys, accumulating sulphide talus from the breakdown of unstable chimneys, and from minerals precipitating inside the mounds^{3,4}. This is in contrast to smaller, immature deposits found elsewhere on the sea floor⁷. Local supergene enrichment observed on the scale of individual grains in samples from the TAG field resembles larger scale processes in Cu-Fe-sulphide deposits on land. Oxidized ores in ancient massive sulphides are commonly enriched in gold as finely divided microscopic grains in the zone of secondary Cu-sulphides (chalcocite, digenite, and covellite) beneath the water table². Likewise, native copper is commonly found in the siliceous, limonitic gossans overlying the altered primary sulphides¹⁰. The degree of enrichment is sometimes spectacular; the Mt Morgan massive sulphide deposit in Queensland, Australia, for example, was capped by 2.4 million tonnes of oxidized ore with an average grade of 30.6 p.p.m. Au (refs. 13, 14). Supergene enrichment of gold and copper in these ancient massive sulphides has previously been attributed to oxidation by near-surface ground-

water following exposure by uplift and erosion². The enrichment of gold in some iron-oxide gossans has also been attributed to the removal of a large volume of gangue minerals by differential leaching (such as in the oxidized caps of ophiolite-hosted massive sulphide deposits in Cyprus)^{15,16}. But processes involving chemical transport and redeposition of gold and copper, like those observed in the TAG field, may have been responsible for supergene enrichment of many ancient massive sulphides while they were still on the sea floor. Thus, supergene gold and copper may exist undiscovered in massive sulphide deposits that have never been exposed to near-surface weathering on land. Although it is premature to comment on the economic viability of modern sea floor massive sulphides, the discovery of supergene enrichment of gold and copper in the TAG field may be an important factor in the future economic consideration of sea floor deposits.

We thank the officers and crew of the NOAA ship *Researcher*, the NOAA Vents Program, and the NOAA Office of Undersea Research for help in obtaining samples for this study. R. V. Hancock (Slowpoke Nuclear Reactor, University of Toronto) and M. P. Gorton greatly assisted the neutron activation analyses. Our work on gold in massive sulphides is made possible by the Natural Sciences and Engineering Research Council of Canada and Energy, Mines and Resources, Canada. This paper benefitted from discussions with T. F. McConachy and P. M. Herzig. P. B. Barton Jr is thanked for his many helpful comments and suggestions on an early draft of the paper.

Received 11 January 1988; accepted 4 March 1988.

1. Rona, P. A., Klinkhammer, G., Nelson, T. A., Trefry, J. H. & Elderfield, H. *Nature* **321**, 33-37 (1986).
2. Boyle, R. W. *Geol. Soc. Can. Bull.* **280**, 584 (1979).
3. Rona, P. A., Pockalny, R. A. & Thompson, G. *Am. Geophys. Union Trans. Abstr.* **67**, 1021 (1986).
4. Thompson, G., Humphris, S. E., Schroeder, B., Sulanowska, M. & Rona, P. A. *Can. Mineralogist* **26**, (in the press).
5. Thompson, G., Humphris, S. E. & Rona, P. A. *Geol. Soc. Amer. Abstr.* **18**, 772 (1986).
6. Schroeder, B., Thompson, G., Humphris, S. E., Sulanowska, M. & Rona, P. A. *EOS* **67**, 1022 (1986).
7. Hannington, M. D., Peter, J. M. & Scott, S. D. *Econ. Geol.* **81**, 1867-1883 (1986).
8. Andrew, R. L. *Miner. Sci. Engng* **12**, 193-215 (1980).
9. Blain, C. F. & Andrew, R. L. *Miner. Sci. Engng* **9**, 119-150 (1977).
10. Thornber, M. R. *Chem. Geol.* **53**, 279-301 (1985).
11. Mann, A. W. *Econ. Geol.* **79**, 38-49 (1984).
12. Webster, J. G. *Geochim. cosmochim. Acta* **50**, 1837-1845 (1986).
13. Cornelius, K. D. *Econ. Geol.* **64**, 885-902 (1969).
14. Taube, A. *Econ. Geol.* **81**, 1322-1340 (1986).
15. Bear, L. M. *Cyprus Geol. Surv. Bull.* **1**, 208 (1963).
16. Bischoff, J. M., Rosenbauer, R. J., Aruscavage, P. J., Baedeker, P. A. & Crock, J. G. *Econ. Geol.* **78**, 1711-1720 (1983).

Bacteria and cyanobacteria associated with phytodetritus in the deep sea

K. Lochte* & C. M. Turley†

* Institut für Meereskunde an der Universität Kiel, Düsternbrooker Weg 20, D-2300 Kiel, FRG

† Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

Sedimentation of phytodetrital aggregates from euphotic waters to the deep sea has been recognized in recent years as probably the most important source of energy for the deep-sea ecosystem¹⁻³. Little is known, however, of the processes by which the detritus is degraded. We report here on phytodetrital aggregates collected from the sediment surface 4,500 m deep in a mid-oceanic area of the northeastern Atlantic. The aggregates contained a rich community of active bacteria and cyanobacteria. The abundance of intact cyanobacteria indicates that the phytodetritus is sedimented rapidly and that picoplankton, which form a significant fraction of primary production during summer months, can be transported from surface waters to the deep sea by attachment to aggregates. Rapid bacterial growth occurred on the detrital material when incubated under deep-sea conditions and similar growth rates were found under surface-water incubation conditions. The results show that phytodetritus is rapidly used by deep-sea adapted bacterial populations and that the biological degradation and transformation of sedimented detrital material in the deep sea is faster than hitherto assumed.

The deposition of detrital material within a short period of time after the spring phytoplankton bloom^{4,5} represents a major nutrient input to the deep-sea community, provided its organic content can be readily used by the deep-sea biota. A close coupling of sedimentation and microbial breakdown of the deposited detrital material has been observed in shallow benthic ecosystems⁶, but there are no comparable data for the deep sea. Although bacterial transformation of organic matter in the deep sea has been considered in the past to be an extremely slow process⁷, evidence has been found recently of rapid microbial action at abyssal depths^{8,9}. Our aim therefore was to observe the rate of microbial degradation of naturally occurring detrital material collected from abyssal depths under simulated deep-sea and shallow-water conditions in order to compare bacterial activity in the two regimes.

During this investigation phytodetritus was found, for the first time, on the sediment surface at abyssal depths in a mid-oceanic area (Table 1) in the northeastern Atlantic¹⁰, where influences of coastal areas or the shelf break cannot be expected. Immediately after sampling, the phytodetritus was incubated on board ship under simulated deep-sea (450 atm, 2 °C) or surface-water (1 atm, 15 °C) conditions and changes in the microbial populations, particulate organic carbon (POC), particulate organic nitrogen (PON) and C/N ratios monitored (for details, see legends to Figs 1 and 2). In addition, microbiological and chemical composition of different phytodetrital samples from the same area were analysed (Table 1).

Immediate microscopic observation of 'greenish' phytodetrital samples revealed intact small chlorophytes, coccolithophorids and large numbers of strongly autofluorescent cyanobacterial cells, many of which were dividing, resembling cells of the genus *Synechococcus*^{11,12}. During incubation under both regimes, an increase in cyanobacteria was observed for the first day (Fig. 1b, Table 2), and may or may not have been a response to their short (1-2 h) exposure to light during sampling. Their subsequent decrease could be due to autolysis or grazing. The peak in PON content (Fig. 2b) coincides with the high occurrence of cyanobacteria (Fig. 1b). As the pigment phycoerythrin may act as a reservoir of amino nitrogen¹³, the strong

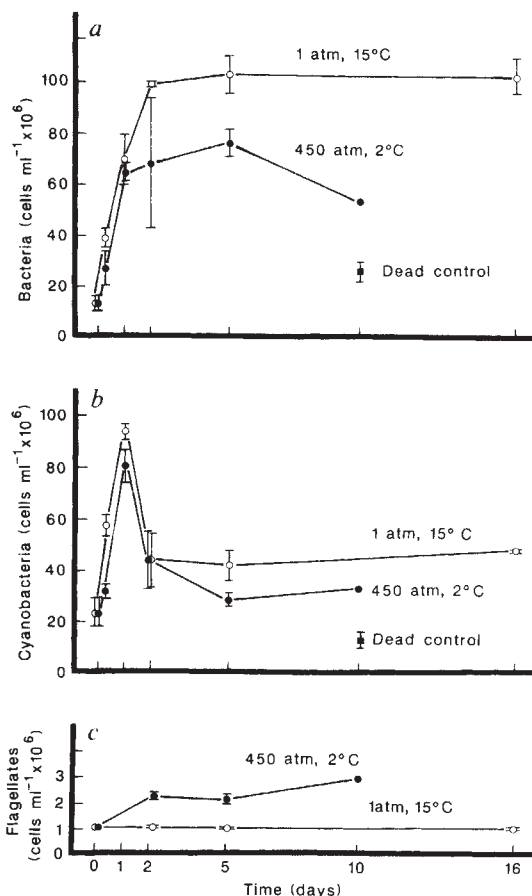


Fig. 1 Changes in natural populations of *a*, bacteria *b*, cyanobacteria and *c*, flagellates on deep-sea phytodetritus sampled on the 12th August (see Table 1) under simulated deep-sea (450 atm, 2 °C) and surface-water (1 atm, 15 °C) conditions. After sampling the sediment and overlying water using a multiple corer¹⁵, the whole phytodetrital layer was immediately aseptically removed by wide-mouthed pipette from the surface of 12 undisturbed sediment cores¹⁵, pooled, dispensed into sterilized polyethylene bags and placed in pressure vessels either at 450 atm and 2 °C or at 1 atm, 15 °C in the dark. The dead control received 2% formalin (final volume). Each sampling point represents the mean of two replicate incubations and the bars the range of values. Bacterial¹⁶ and flagellate¹⁸ numbers were determined by acridine orange epifluorescence direct counts. Cyanobacteria were counted directly by epifluorescence under green excitation (excitation filter 515-560 nm, barrier filter 580 nm), under which the phycoerythrin-containing *Synechococcus* cyanobacteria autofluoresces orange-red¹¹.

phycoerythrin fluorescence in the cyanobacteria on the phytodetritus may therefore be an indication of high nitrogen content, rather than of an active photosynthetic system.

Cyanobacteria have been found in bathypelagic marine snow particles^{12,14} and in sediment traps at 2,000 m (ref. 14). Owing to their small size, their own sinking rate is negligible and their occurrence in the deep sea can only be explained by rapid transport through the water column attached to larger particles. *Synechococcus* and *Synechocystis* cyanobacteria are important primary producers, especially in oceanic waters during summer^{11,13}. The surface waters at the study site during this investigation contained 5×10^5 cyanobacteria per ml, sediment contact water at 4,500 m contained 3×10^2 cyanobacteria per ml, and the phytodetritus samples contained $8-20 \times 10^6$ cyanobacteria per ml. Cyanobacteria were also abundant in detritus collected from a further four multiple sediment corers¹⁵ in this region; a polychaete tube; a vertical plankton net haul through the whole water column, and two closing net hauls from 150 m and 350 m

Table 1 Chemical and microbiological observations of deep-sea benthic phytodetritus sampled by multiple sediment corer¹⁵ from different stations in the NE Atlantic during July/August 1986

Sampling date	Station position		Sample depth (m)	Chemical observations		Microbiological observations				
	Latitude (°N)	Longitude (°W)		Dry weight (mg ml ⁻¹)	POC ²³ (µg C mg ⁻¹)	Cell numbers (×10 ⁶ mg ⁻¹)	Mean cell volume (µm ³)	Biomass (µg C mg ⁻¹)	Biomass (as % of POC)	FDC (%)
28 July	47°26.20'	19°43.08'	4,552	2.03	78.5	165.96	0.204	3.72	4.7	3.7
28 July	47°26.33'	19°42.99'	4,554	36.20	8.8	6.03	0.137	0.09	1.0	6.1
2 August	47°25.42'	19°42.05'	4,551	9.20	10.1	9.36	NA	NA	NA	1.0
4 August	47°20.50'	19°43.40'	4,157	21.01	18.6	12.46	0.150	0.21	1.1	7.0
7 August	47°24.71'	19°39.91'	4,536	3.55	20.1	39.04	0.112	0.48	2.4	2.0
7 August	47°25.35'	19°39.46'	4,508	11.46	14.4	21.20	0.158	0.37	2.6	3.8
12 August	47°24.19'	19°46.46'	4,546	3.54	19.1	4.72	0.149	0.08	0.4	NA
12 August	47°24.19'	19°46.46'	4,546	3.17	26.5	3.28	0.152	0.06	0.2	4.0
Cyanobacteria†										
7 August	47°24.71'	19°39.91'	4,536	3.55	20.1	2.83	0.288	0.09	0.4	6.0
7 August	47°24.71'	19°39.91'	4,536	3.55	20.1	3.70	0.288	0.12	0.6	10.1
12 August	47°24.19'	19°46.46'	4,546	3.54	19.1	2.23	0.288	0.07	0.3	10.2
12 August	47°24.19'	19°46.46'	4,546	3.17	26.5	2.35	0.288	0.07	0.3	7.3

* Bacterial cell numbers, frequency of dividing cells (FDC), mean cell volume and biomass were determined by direct counts using acridine orange epifluorescence¹⁶.

† Cyanobacterial numbers¹¹ and FDC were estimated on counts >500 cells, mean cell volume was calculated from the measurements of 200 cells and cyanobacterial biomass calculated from total cell volume using the volume to carbon conversion factor of 1.06×10^{-13} g C µm⁻³ (ref. 24). Phytodetrital samples for enumeration of cyanobacteria were stored in 2% buffered formalin, at 5 °C, in the dark. Strong autofluorescence of the cyanobacteria even remained 8 weeks after sampling. NA, data not available.

water depth. This indicates that their association with aggregates occurred through the whole water column, and this rapid transport of photosynthetic picoplankton to the deep-sea floor was an ongoing process during the period of investigation. Furthermore, the presence of 'whitish', more-degraded detrital aggregates on the sea floor, composed of diatoms characteristic of the spring bloom (unpublished observation), indicates that sedimentation of aggregates may occur on several occasions during late spring and summer in these waters.

An abundant heterotrophic bacterial assemblage colonized the phytodetritus consisting of bacteria considerably larger (Table 1) than planktonic bacteria from the surface waters¹⁶. Phytodetrital age, degree of decomposition and sampling are likely to contribute to the tenfold variations in chemical and biological measurements (Table 1). Incubation under deep-sea conditions resulted in rapid bacterial growth similar to that under surface-water conditions during the first two days of incubation (Fig. 1a, Table 2). High growth rates are also likely to occur in other phytodetrital samples as indicated by the relatively high bacterial FDC (frequency of dividing cells) values (Table 1). Indeed, the bacterial growth rates found under pressure here (Table 2) and by Deming⁸ are as fast as those found in surface waters elsewhere¹⁷.

A flagellate population developed only under pressure (Fig. 1c), indicating that this organism was barophilic. This flagellate resembled the barophilic, bacterivorous flagellate of the genus *Bodo* isolated from these waters¹⁸. The flagellate growth rate (Table 2) was comparable to the lower growth rates determined on shallow-water flagellates¹⁹. Grazing of bacteria by this flagellate is probably the cause of the lower yield of bacteria in the pressure incubations (Fig. 1a).

During the first 49 h, the POC concentration rose by 17.1 and 15.5 µg C ml⁻¹ in the simulated deep-sea and surface-water incubations respectively (Fig. 2a). This increase occurs when bacteria and cyanobacteria grow (Fig. 1a and b), and represents 92% and 84% of the dissolved organic carbon²⁰ (DOC) determined in the interstitial water of phytodetrital samples (average 18.35 µg C ml⁻¹, mean of 18.30 and 18.40 µg C ml⁻¹). This concentration was eightfold higher than that measured in the surrounding water (2.31 µg C ml⁻¹, mean of 2.21, 2.37 and 2.34 µg C ml⁻¹). It would seem that the interstitial DOC is rapidly converted into bacterial biomass during the initial incubation period. Subsequently, the POC concentration gradually decreases at a rate of 1.49 µg C ml⁻¹ per day and 2.23 µg C ml⁻¹ per day, which is equivalent to a POC decomposition of 1.8 and

2.9% per day in the deep-sea and surface-water incubations respectively. The loss of POC is caused by microbial respiration during their decomposition of the detritus. This is substantiated by there being no loss in PON (Fig. 2b), and a gradual decrease in the C/N ratio of the detritus towards the end of the incubations (Fig. 2c). This transformation of detrital carbon into bacterial carbon may be important to organisms feeding on the detritus, such as small flagellates¹⁸, benthic foraminiferans²¹ and holothurians²².

The discovery of dividing cyanobacteria in the deep sea indicates that the transport of recently fixed primary production by sedimenting aggregates is a relatively rapid process. This, the fast bacterial and flagellate growth on the phytodetritus under deep-sea conditions, and their subsequent decomposition and transformation of the aggregates indicate that sedimentation of phytodetrital aggregates could be of greater significance in the material flow to the deep sea than hitherto supposed.

We thank Frauke Dreyer for technical assistance, A. V. Altenbach for help with organic carbon determinations, F. Riemann

Table 2 Mean relative growth rates (μ_n) and doubling times (D_n) for bacteria, cyanobacteria and flagellates growing on deep-sea phytodetritus incubated under simulated deep-sea and surface-water conditions

Organisms	$\log_e N_{t1}$ (cells ml ⁻¹)	$\log_e N_{t2}$ (cells ml ⁻¹)	$t_2 - t_1$ (days)	μ_n (per day)	D_n (days)
Simulated deep-sea conditions (450 atm, 2 °C)					
bacteria	16.42	17.99	1.08	1.45	0.48
cyanobacteria	16.95	18.20	1.08	1.16	0.60
flagellates	10.60	14.05	2.04	1.69	0.41
Simulated shallow water conditions (1 atm, 15 °C)					
bacteria	16.42	18.06	1.08	1.52	0.46
cyanobacteria	16.95	18.36	1.08	1.31	0.53
flagellates	10.60	10.60	2.04	0.00	0.00

See legend to Fig. 1 for further details on sampling and incubation. $\mu_n = (\log_e N_{t2} - \log_e N_{t1}) / (t_2 - t_1)$, where N_{t1} is the number of cells at the beginning of the experiment (t_1), N_{t2} is the number of cells at the end of the experimental growth phase (t_2), and D_n was calculated by $0.693 / \mu_n$.

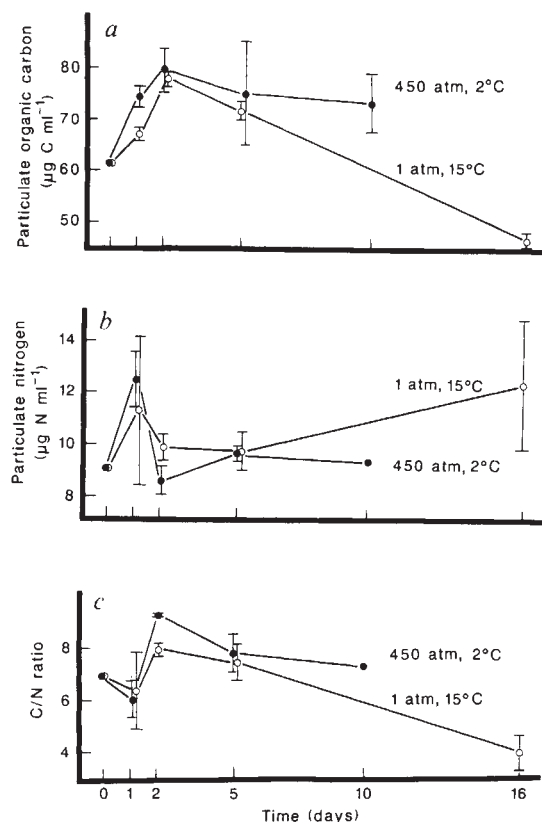


Fig. 2 Changes in *a*, particulate organic carbon determined by wet oxidation²³, *b*, particulate organic nitrogen determined by elemental CHN analysis²⁵ and *c*, particulate organic carbon: particulate organic nitrogen ratio (C/N) of the deep-sea phytodetritus incubated under simulated deep-sea (450 atm, 2°C) and surface-water (1 atm, 15°C) conditions (see legend to Fig. 1 for sampling and incubation details).

for making his multiple corer and net haul samples available, scientists and crew on cruise *METEOR 3* for help in obtaining samples. K.L. was supported by the Minister für Forschung und Technologie der Bundesrepublik Deutschland as part of the BIOTRANS programme. C.M.T. acknowledges support from the Royal Society, a John Murray Travel Scholarship and a Browne-Hill-Marshall Grant.

Received 9 December 1987; accepted 23 March 1988.

1. Billett, D. S. M., Lampitt, R. S., Rice, A. L. & Mantoura, R. F. C. *Nature* **302**, 520–522 (1983).
2. Honjo, S. *Science* **218**, 883–884 (1982).
3. Fowler, S. W. & Knauer, G. A. *Prog. Oceanogr.* **16**, 147–194 (1986); **33**, 70–73 (1988).
4. Lampitt, R. S. *Deep Sea Res.* **32**, 885–897 (1985).
5. Rice, A. L. *et al. Proc. R. Soc. Edinb.* **88B**, 265–279 (1986).
6. Graf, G., Bengtson, W., Diesner, U., Schulz, R. & Theede, H. *Mar. Biol.* **67**, 201–208 (1982).
7. Jannasch, H. W. & Wirsen, C. O. *Science* **180**, 641–643 (1973).
8. Deming, J. W. *Mar. Ecol. Prog. Ser.* **25**, 305–312 (1985).
9. Cole, J. J., Honjo, S. & Erez, J. *Nature* **327**, 703–704 (1987).
10. Cruise Report of Research Vessel *METEOR*, cruise No. 3, leg *BIOTRANS IV* (Abteilung Hydrobiologie, Zeiseweg 9, D-2000 Hamburg 50, FRG, 1986).
11. Johnson, P. W. & Sieburth, J. McN. *Limnol. Oceanogr.* **24**, 928–935 (1979).
12. Silver, M. W. & Alldredge, A. L. *J. Mar. Res.* **39**, 501–530 (1981).
13. Fogg, G. E. *Proc. R. Soc. B228*, 1–30 (1986).
14. Silver, M. W., Gowing, M. M. & Davoll, P. J. *Can. Bull. Fish. Aqu. Sci.* **214**, 311–341 (1986).
15. Barnett, P. R. O., Watson, J. & Connelly, D. *Oceanologica Acta* **7**, 399–408 (1984).
16. Lochte, K. & Pfannkuche, O. *Mar. Ecol. Prog. Ser.* **39**, 153–164 (1987).
17. Ducklow, H. W. *Bioscience* **33**, 494–501 (1983).
18. Turley, C. M., Lochte, K. & Patterson, D. J. *Deep Sea Res.* (in the press).
19. Fenchel, T. *Mar. Ecol. Prog. Ser.* **8**, 225–231 (1982).
20. Mantoura, R. F. C. & Llewellyn, C. A. *Geochim. cosmochim. Acta* **47**, 1293–1309 (1983).
21. Gooday, A. *Nature* **332**, 70–73 (1988).
22. Billett, D. S. M., Llewellyn, C. & Watson, J. *Proc. 6th Int. Echinoderm Conf.* (ed. Burke, R. D.) (in the press).
23. Altenbach, A. V. *J. foram. Res.* **17**, 106–109 (1987).
24. Nagata, T. *Appl. envir. Microbiol.* **52**, 28–32 (1986).
25. Newell, R. C., Lucas, M. I. & Linley, E. A. S. *Mar. Ecol. Prog. Ser.* **6**, 123–136 (1981).

A fitness cost of foraging in the guppy

Jean-Guy J. Godin* & Shelley A. Smith

Department of Biology, Mount Allison University, Sackville, New Brunswick, Canada E0A 3C0

One hypothesized cost of foraging is an increased risk of mortality due to predation, which may result from greater conspicuousness and reduced vigilance for predators^{1–5}. At high prey densities, a foraging animal must compromise vigilance to overcome the 'confusion effect'^{6,7} and achieve high feeding rates^{8,9}. The optimal tradeoff between food intake rate and predator avoidance should depend on the forager's current hunger or energy deficit, as risk of starvation is assumed to be dependent on hunger state^{8,10–12}. Therefore, the probability of a foraging animal being captured by a predator is expected to increase with its food deficit and food density, and thus be dependent on its current feeding rate^{1,8–12}. But present evidence for an increased risk of predation associated with foraging is either circumstantial^{3,4,13–16}, or indirect⁹. We report here the first direct evidence for a fitness-associated cost of foraging which is dependent on the forager's current feeding rate, its food (energy) deficit and local food density. Female guppies, *Poecilia reticulata*, foraging on zooplankton achieved higher feeding rates and were increasingly more likely to be captured by an ambush predator as their food deficit and prey density increased, and the pre-attack feeding rate of guppies which were captured was higher on average than that of survivors.

To investigate the fitness cost of foraging we used a laboratory model predator-prey system consisting of individual female guppies foraging on live water-fleas (*Daphnia magna*), at different densities, under threat of predation from a live ambush predator, the jewel cichlid fish (*Hemichromis bimaculatus*). Cichlid fishes are natural predators of the guppy¹⁷. Our guppies were obtained from a commercial supplier and were of unknown origin. They were of similar length (22.6 ± 0.2 mm, mean $\pm$ s.e.m.), weight (0.13 ± 0.04 g) and colour, and none were pregnant when tested. The fish were held in a communal aquarium maintained at 22–24 °C, and fed twice daily with laboratory-cultured *Daphnia* and a flake food (TetraMin) for at least 2 weeks before testing. Only *Daphnia* passing through a sieve with 1.0-mm holes but not through one with 0.5-mm holes, carrying no eggs and uniformly coloured were selected for the experiments, so as not to reduce the potential confusion effect. Four aquarium-reared cichlid fish (9.7 ± 0.1 cm, 17.9 ± 0.6 g) were used as predators. Each cichlid was maintained separately in identical test aquaria, wherein they readily attacked guppies when given the opportunity.

Four identical Plexiglas test aquaria (78 × 22 × 25 cm) had three sides covered externally with light blue styrofoam and an observation blind mounted in front of the fourth side. The aquaria had gravel on their bottom, contained continually filtered water maintained at 22–24 °C and were illuminated by overhead fluorescent lights (830 Lx) on a 16-h light:8-h dark cycle. Each aquarium was divided into two compartments by a white plastic screen partition, which could be raised remotely. The smaller compartment (18 × 22 × 25 cm) contained a plastic plant anchored to the substrate, received less light (500 Lx) and housed a cichlid fish predator. The larger compartment had no plants and served as a foraging arena for the test guppy. Based on the behaviour of the fish, the guppy apparently could not see the cichlid lying in ambush and in dimmer light near the submerged plant behind the screen partition, but the cichlid seemed able to see the guppy. An attack on the nearby guppy was staged by raising the partition.

* Present address: Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

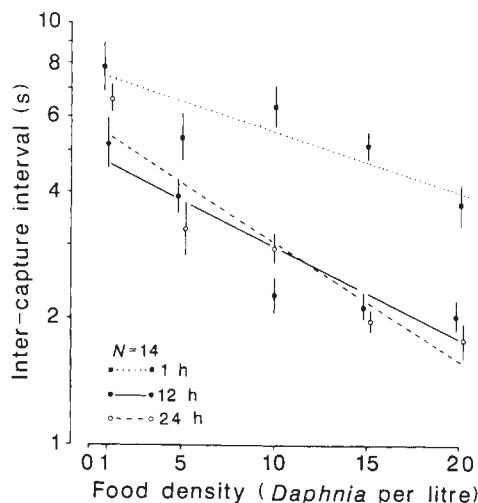


Fig. 1 Observed inverse relationship between *Daphnia* density and instantaneous feeding rate of guppies, expressed as mean ($\pm$ s.e.m.) $\log_{10}$ interval between six successive *Daphnia* captures, is significant at each of three food deprivation levels (regression coefficients significantly different from zero, all $P < 0.001$). Guppy feeding rate therefore increased significantly, but at a diminishing rate, with increasing food density. The slopes of the three regression lines are significantly different from each other ($F = 5.3$, d.f. = 2,204, $P < 0.01$, analysis of co-variance (ANCOVA)); this is attributed mainly to the regression coefficients for the 1- and 24-h food deprivation levels being different from one another ($P < 0.05$, Tukey-Kramer test). The slopes and Y -intercepts of the relationships obtained for the 12- and 24-h food-deprived guppies were not significantly different from one another ($P > 0.05$, Tukey-Kramer tests), whereas both these Y -intercepts were different (both $P < 0.05$, Tukey-Kramer tests) from the one obtained for the relationship at the 1-h deprivation level. Therefore, guppies which were previously deprived of food for 12 or 24 h responded similarly to increasing *Daphnia* density, and both groups fed at significantly higher rates than the 1-h food-deprived guppies. Key: ■, 1 h; ●, 12 h; ○, 24 h. $N = 14$.

In the first series of experiments, we quantified guppy feeding rates in the absence of predation hazard (cichlid not present in the aquarium) at each of five food densities (1, 5, 10, 15 and 20 *Daphnia* per litre) combined with each of three prior food deprivation (deficit) treatments (1-, 12- and 24-h periods). Each combination of food density and food deficit level was replicated 14 times in random order with different guppies, for a total of 210 trials. Each trial consisted of placing a single guppy, which had previously been deprived of food for either 1, 12 or 24 h after having been fed five *Daphnia*, in the larger aquarium compartment and allowing it to acclimatize for 1 h. A predetermined number of *Daphnia* in 30 ml of water were then introduced into the centre of the compartment without disturbing the fish; the *Daphnia* quickly dispersed within the water column. The time elapsed between each of the first six *Daphnia* eaten by the guppy was measured using an audio recorder and digital stopwatch. The five successive inter-capture intervals thus obtained were normalized by logarithmic transformation and averaged for each guppy. The reciprocal of this value represents the fish's instantaneous feeding rate. Satiation effects on feeding rate were avoided by limiting the guppy's intake to only six *Daphnia* (<15% of stomach capacity).

Guppy food intake increased at a diminishing rate with increasing *Daphnia* density, and intake rate was greater at the higher food deficit levels. This is represented by significant inverse log-linear relationships between food density and mean inter-capture interval for all three food deficit levels (Fig. 1). Both food density ($F = 40.5$, d.f. = 4, 195, $P < 0.001$, two-way analysis of variance (ANOVA)) and food deficit level ($F = 61.9$, d.f. = 2, 195, $P < 0.001$) significantly increased feeding rate, and the effects of these two variables were not independent of one

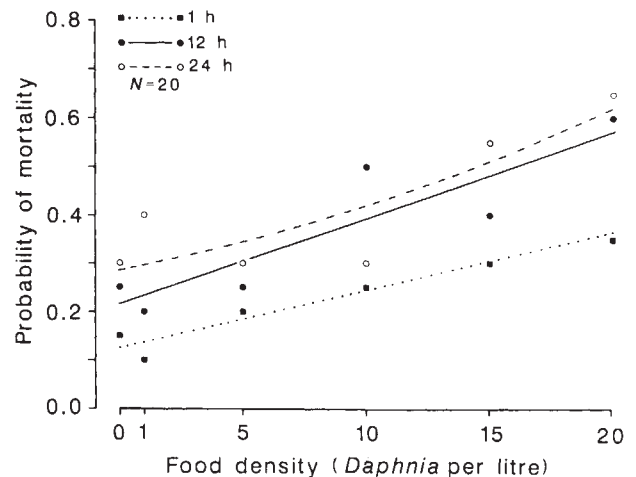


Fig. 2 The probability of an individual guppy being captured by a cichlid predator on a given attack is a positive function of local food density. Risk of mortality to predation is also dependent on the forager's hunger level, with the 12- and 24-h food-deprived guppies incurring greater risks than the fish with the lowest hunger level, at any given food density. To illustrate trends, least-squares regression lines were plotted separately for each food deprivation level. Key: ■, 1 h; ●, 12 h; ○, 24 h. $N = 20$.

another (interaction term, $F = 3.5$, d.f. = 8, 195, $P < 0.001$). An asymptotic feeding rate (of ~ 1 *Daphnia* per 2 s) was reached at the highest two food densities for guppies with 12- and 24-h food deficits; the 1-h food-deprived fish did not attain such an asymptote over the range of food densities used (Fig. 1). The feeding-gain functions for the 12- and 24-h food-deprived guppies were not significantly different from one another, but both were different from that of fish at the 1-h food deficit level (Fig. 1).

On the basis of these observations, we predicted that guppies would incur a progressively greater individual risk of mortality to predation with increasing *Daphnia* density at each of the food deficit levels. This risk is expected to be rather similar for the 12- and 24-h food-deprived fish, but both these groups should suffer greater risks than the 1-h food-deprived fish which fed at much lower rates. To test these predictions we used the same experimental apparatus, protocol and design as before, with the following exceptions. A cichlid predator was now present in the smaller compartment of the aquarium, a control food treatment (0 *Daphnia* per litre) was added to the other five food densities and 20 different guppies were tested at each combination of food density and food deficit level, for a total of 360 randomized trials. In each trial a single guppy was allowed to eat at least two *Daphnia* before the screen partition was slowly raised 10 cm, following which the cichlid predator attacked the foraging guppy within 11.8 ± 0.7 s on average. Predator attack latency did not differ between food density treatments ($F = 1.9$, d.f. = 5, 354, $P > 0.1$). The interval between each food item eaten by the guppy before the attack and whether it was captured or escaped the initial attack were recorded. The cichlid was allowed to eat the guppy, even if its initial attack was unsuccessful, so as to standardize its hunger (one guppy per day) between daily trials. At the end of each trial, the remaining *Daphnia* were removed from the aquarium. Probability of mortality was calculated as the ratio of the number of guppies captured on the initial attack to the number (20) independently tested for each treatment combination. Only the initial attack is relevant here, as an unsuccessful attack in nature would probably result in the prey finding nearby refuge and thus avoiding further immediate attacks.

Probability of mortality to predation was significantly increased by food density ($G = 32.1$, d.f. = 15, $P < 0.005$) and,

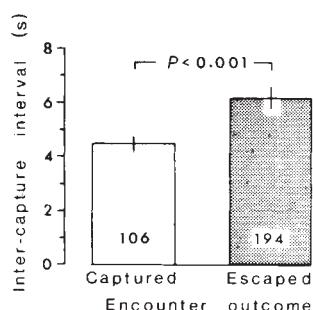


Fig. 3 Guppies captured by the cichlid predator on its initial attack had on average a significantly greater ($t = 3.3$, d.f. = 298, $P < 0.001$) instantaneous feeding rate, immediately before the attack, than guppies which escaped the attack. Feeding rate is expressed as the mean ($\pm$ s.e.m.) interval between the penultimate and last *Daphnia* eaten by the guppy before the predator's attack. The means are based on $\log_{10}$ -transformed data, pooled over all treatment combinations for each category of guppies (captured versus escaped), respectively. The number within the bars denotes the replicate number of guppies.

less dramatically, by food deficit level ($G = 14.3$, d.f. = 6, $P < 0.05$), as revealed by a three-way analysis of independence using a log-linear model (G -test) for frequency data¹⁸. For each food deficit level, probability of mortality was an increasing function of food density, as predicted (Fig. 2). This relationship is significant ($G = 11.2$, d.f. = 5, $P < 0.05$) for the 12-h food-deprived fish but not so at the other food deficit levels, although it approached significance at the 24-h food deficit level ($G = 9.5$, d.f. = 5, $P > 0.05$). Nevertheless, guppies feeding at the highest food density (Fig. 2) suffered a 120–140% greater risk of mortality to predation, depending on their food deficit, compared with non-feeding guppies (at 0 *Daphnia* per litre); this increase in risk is significant at the two highest food deficit levels (both $P < 0.05$, d.f. = 1, G -tests). As expected from the results of the first series of experiments on feeding rates, risk of mortality was more similar between the 12- and 24-h food-deprived guppies than between either of these groups and the 1-h food-deprived fish (Fig. 2). Because the guppies only had time to eat on average between 3.3 (at 1 *Daphnia* per litre) and 3.9 *Daphnia* (at the highest prey density; overall range = 2–7) before the predator's attack, which represents ~8–10% of their stomach capacity, satiation could not have confounded the instantaneous feeding rates and risks of mortality to predation observed.

To test whether the guppy's pre-attack instantaneous feeding rate influenced the outcome of its encounter with the predator, we compared the last inter-capture interval before the initial attack, averaged over all treatment combinations, for guppies which subsequently escaped the attack with those which were captured. Guppies captured on the initial attack had on average a significantly greater instantaneous feeding rate immediately before the attack than fish which escaped the attack (Fig. 3). This is as expected if a high feeding rate constrains vigilance for predators and reduces the likelihood of predator detection⁹.

This study has demonstrated a direct and realized fitness cost of foraging, in the form of an increased risk of mortality to predation, which is dependent on the forager's current feeding rate, its food deficit and local food density. Hitherto the best evidence suggestive of such a foraging cost was Milinski's⁹ observation that stickleback fish feeding at high rates on dense prey are less likely to detect an overhead cryptic model predator than fish feeding on less dense prey. Although reproductive fitness (fecundity) and the likelihood of avoiding starvation are dependent on food intake rate in many species^{19–22}, foragers which attempt to maximize this currency will necessarily incur a greater risk of mortality to predation, as shown here. Because of this cost of foraging, animals are expected to trade off foraging gains for greater safety from predators in such a way as to

maximize fitness. Individuals can achieve such a compromise by shifting habitats, reducing locomotor activity, altering feeding rates, increasing the level of vigilance or joining social groups, for example, when risk of predation is perceived to be high^{3–5,8,10,13–16,23,24}. Such behavioural flexibility is adaptive when risk of predation is dependent on feeding rate, hunger and food density.

We thank L. Taylor for technical assistance, A. Burt, P. Harvey, A. I. Houston and J. R. Krebs for their comments on an earlier version of the paper, and NSERC of Canada and the Donner Canadian Foundation for support (grants to J.-G.J.G.).

Received 8 March 1988; accepted 28 March 1988.

- Krebs, J. R. *Ardea* **68**, 83–90 (1980).
- Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton University Press, 1986).
- Milinski, M. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 236–252 (Croom Helm, Princeton, London, 1986).
- Dill, L. M. *Can. J. Zool.* **65**, 803–811 (1987).
- Lima, S. L. *J. theor. Biol.* **124**, 303–316 (1987).
- Milinski, M. *Z. Tierpsychol.* **43**, 311–325 (1977).
- Ohguchi, O. *Z. Tierpsychol. Suppl.* **23**, 1–79 (1981).
- Heller, R. & Milinski, M. *Anim. Behav.* **27**, 1127–1141 (1979).
- Milinski, M. *Anim. Behav.* **32**, 1157–1162 (1984).
- Milinski, M. & Heller, R. *Nature* **275**, 642–644 (1978).
- Mangel, M. & Clark, C. W. *Ecology* **67**, 1127–1138 (1986).
- McNamara, J. M. & Houston, A. I. *Am. Nat.* **127**, 358–378 (1986).
- Caraco, T., Martindale, S. & Pulliam, H. R. *Nature* **285**, 400–401 (1980).
- Lima, S. L. *Oecologia* **66**, 60–67 (1985).
- Fraser, D. F. & Huntingford, F. A. *Ethology* **73**, 56–68 (1986).
- Fraser, D. F. & Gilliam, J. F. *Behav. Ecol. Sociobiol.* **21**, 203–209 (1987).
- Liley, N. R. & Seghers, B. H. in *Function and Evolution in Behaviour* (eds Baerends, G., Beer, C. & Manning, A.) 92–118 (Clarendon, Oxford, 1975).
- Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (Freeman, New York, 1981).
- Schoener, T. W. A. *Rev. Ecol. Syst.* **2**, 369–404 (1971).
- Wootton, R. J. *Symp. zool. Soc. Lond.* **44**, 133–159 (1979).
- Kettersen, E. D. & Nolan, V., Jr. *Ecology* **57**, 679–693 (1976).
- Real, L. A. & Caraco, T. A. *Rev. Ecol. Syst.* **17**, 371–390 (1986).
- Godin, J.-G.J. *Naturaliste can. (Rev. Ecol. Syst.)* **113**, 241–250 (1986).
- Pitcher, T. J. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 294–337 (Croom Helm, London, 1986).

A computational theory for the perception of coherent visual motion

Alan L. Yuille* & Norberto M. Grzywacz†‡

* Harvard University Division of Applied Sciences, G12e Pierce Hall, Cambridge, Massachusetts 02138, USA

† Center for Biological Information Processing, Massachusetts Institute of Technology, E25-201, Cambridge, Massachusetts 02139, USA

When we see motion, our perception of how one image feature moves depends on the behaviour of other features nearby. In particular, the Gestaltists proposed the law of shared common fate^{1,2}, in which features tend to be perceived as moving together, that is, coherently. Recent psychophysical findings, such as the cooperativity of the motion system^{3,4} and motion capture^{5,6}, support this law. Computationally, coherence is a sensible assumption, because if two features are close then they probably belong to the same object and thus tend to move together. Moreover, the measurement of local motion may be inaccurate⁷ and so the integration of motion information over large areas may help to improve the performance. Present theories of visual motion, however, do not account fully for these coherent motion percepts. We propose here a theory that does account for these phenomena and also provides a solution to the aperture problem⁸, where the local information in the image flow is insufficient to specify the motion uniquely.

The theory divides the computation of motion into two stages which we refer to as the measuring and the smoothing stages. The measuring stage estimates or restricts the velocity field from the image information. In humans, two processes have been

‡ To whom correspondence should be addressed.

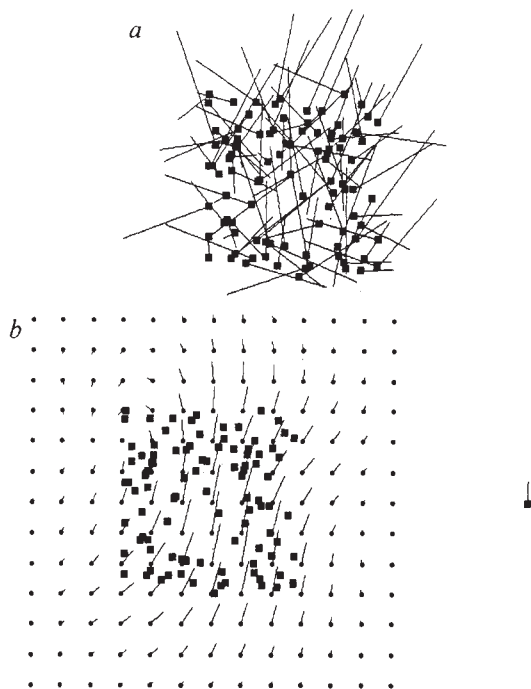


Fig. 1 Motion cooperativity. *a*, One hundred dots were randomly placed inside a square region of unit sides; the dot positions are indicated by the little solid squares. Each of these dots was assumed to be moving. The horizontal components of their velocity were obtained from a homogeneous random distribution whose values ranged from -1 to 1 . The vertical components had the same distribution plus a constant bias upwards of 0.25 . The velocity of each dot is shown by the lines attached to each little square. *b*, The velocity field computed by the motion coherence theory by solving equation (1) with $\lambda = 2.5$ and $\sigma = 0.6$. The computed field is shown as lines attached to little circles positioned in a grid and the position of the moving dots are also shown. (The solution is continuous, however, because it is computed analytically as in equation (3); the grid is shown for purposes of illustration.) The computed field in the region of the moving dots is more coherent than that of the input. Also, the velocity of the computed field is roughly the same as the bias given to the input data. For comparison, an isolated dot moving by the side of the field indicates the bias. This output coherence is similar to the psychophysical phenomenon of motion cooperativity^{3,4}. Another desirable property of the field is that its speed decreases with distance from the moving dots, and thus objects that are not close are guaranteed to be perceived as different objects. The theory in its present form fails for close objects, however, as it does not have any mechanism for detecting motion discontinuities.

suggested to mediate this stage, one that deals with short-range and another that deals with long-range motions⁹. Several schemes have been proposed to measure short-range motion^{8,10-13} and any of them may be used as the theory's measuring stage. For long-range motions the correspondence problem arises¹⁴. Different types of constraint are possible to solve this problem^{14,15}, and each of them predicts different velocity fields. At this stage, our theory merely requires that features correspond, and the exact correspondence is unspecified.

In the smoothing stage, a velocity field is constructed over the entire visual field, even where no estimates of motion have been made. This velocity field is constrained by the restrictions found in the measuring stage and simultaneously is biased to be as smooth as possible. We call this theory the motion coherence theory. (In its present form, the theory does not incorporate mechanisms for motion transparency¹⁶, motion segmentation¹⁷ or motion inertia¹⁸.)

We now present the theory formally, first for short-range motion. Let the velocity measurement at point $\vec{r}_i$ be $M(\vec{V}_i)$,

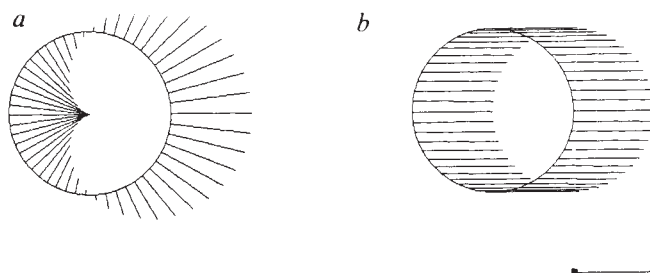


Fig. 2 Aperture problem. A circle of unit radius was assumed to be moving to the right with a velocity of 1.0 . If the motion is measured with local detectors, then only the components of the motion perpendicular to the contour can be measured. This is an example of the aperture problem, and in *a*, these perpendicular components are shown. The velocity field along the contour, as predicted by the motion coherence theory, is shown in *b*. This field was computed by solving equation (1) with $\lambda = 0.001$ and $\sigma = 3$. The isolated dot moving at the bottom of *b* has the correct velocity of the circle, implying that the aperture problem has been solved. The average speed of the computed velocity field along the contour was 0.996 , which is incorrect by only 0.4% .

where $\vec{V}_i$ is the true image velocity. For example, for contour motion, $M(\vec{V}_i)$ would correspond to the perpendicular component of the velocity field or to a rough approximation to it^{8,19}. The theory proposes that the smoothing stage constructs a velocity field, $\vec{v}(\vec{r})$ that minimizes the following functional for both components of $\vec{v}$:

$$E(\vec{v}(\vec{r}), \vec{V}_i) = \sum_i (M(\vec{v}(\vec{r}_i)) - M(\vec{V}_i))^2 + \lambda \int \sum_{m=0}^{\infty} c_m |\nabla^m \vec{v}|^2 \quad (1)$$

where $\lambda \geq 0$ and $c_m \geq 0$ are constants, and ∇^m is the $(m/2)$ th power of the laplacian operator if m is even, and the gradient of the $((m-1)/2)$ th power of the laplacian operator if m is odd. The sum is taken over all points where motion information is available. Minimization of the first term of the equation's right-hand side forces the computed velocity field, $\vec{v}$, to respect the measuring stage. On the other hand, the second term forces field smoothness^{10,19,20}. This term is one of the Tikhonov stabilizers, a class of functionals commonly used in computational vision²¹.

For long-range motion, any choice of feature correspondence determines a measured velocity field. To this field corresponds a new field, $\vec{v}$, for which a functional such as in equation (1) is minimum. Our theory searches among all possible correspondences the one which yields the deepest of these minima. (Not all features are necessarily matched, however, because psychophysics shows that preference may be given to those at large spatial scale²². At least, this preference suggests that a coarse-to-fine heuristic may be used to avoid the combinatorial explosion of considering all possible matches.) Suppose there are a number of points $\vec{r}_a$ at time t and $\vec{r}_i$ at time $t + \delta t$. A matching matrix V_{ai} is defined, so that $V_{ai} = 1$, if the a th point is matched to the i th point, and $V_{ai} = 0$ otherwise. If $V_{ai} = 1$, then the velocity at $\vec{r}_a$ is $\vec{v}(\vec{r}_a) = (\vec{r}_i - \vec{r}_a)/\delta t$. We now minimize the following energy function over the velocity field and all possible matchings that satisfy the cover principle¹⁴.

$$E(\vec{v}(\vec{r}), V_{ai}) = \sum_{ai} V_{ai} \left(\vec{v}(\vec{r}_a) - \frac{(\vec{r}_i - \vec{r}_a)}{\delta t} \right)^2 + \lambda \int \sum_{m=0}^{\infty} c_m |\nabla^m \vec{v}|^2 \quad (2)$$

Minimization of the first term of the equation's right-hand side forces compatibility with the data, while the second term forces smoothness. Ullman's minimal mapping theory¹⁴ is an approximation of the motion coherence theory for large λ (ref. 23).

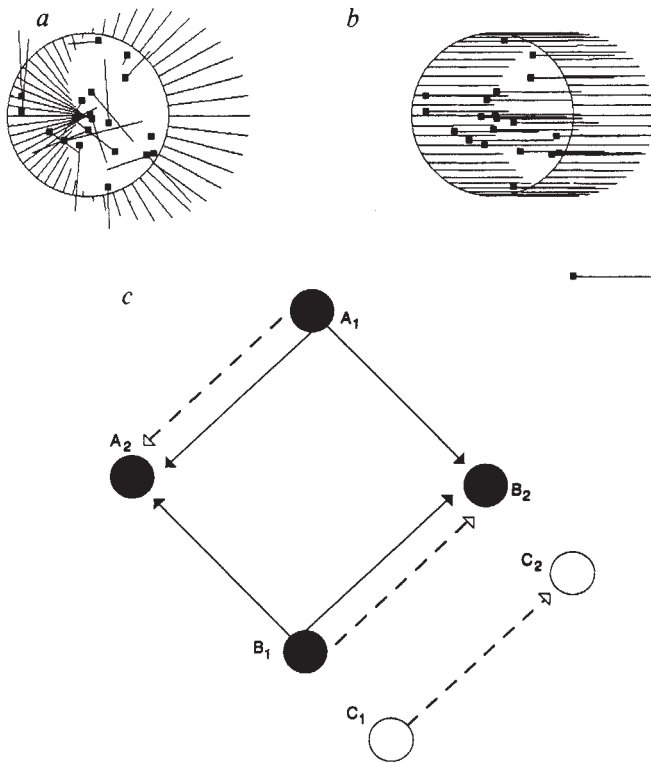


Fig. 3 Motion capture. In *a*, twenty randomly moving dots were placed in random positions inside the circle of Fig. 2. Their motion was random with the same distribution as in Fig. 1, but without the bias. Next, equation (1) was solved with the identical values of λ and σ as in Fig. 2. The result is shown in *b*. As for Fig. 2, the aperture problem was solved for the circle's motion, but this time this solution captured the motion of the internal dots. This result is similar to what humans perceive under the same conditions^{5,6}. The capture predicted by the motion coherence theory can be strong; in the example the average speed of the solution along the contour and over the internal dots was 1.009. The theory also predicts capture for situations where only a few dots move. An example of this is in *c*. Suppose that an apparent motion display is created by alternating two dots between positions A_1 and B_1 , and A_2 and B_2 . If these positions lie on the corners of a square, then an ambiguous motion is perceived³⁰; for example, the motion from B_1 is half the time towards B_2 and half the time towards A_2 . This is represented by the solid arrows. Now, suppose that a third dot is shown in C_1 when the original dots are in A_1 and B_1 , and in C_2 when the original dots are in A_2 and B_2 . According to the theory, due to the spatial proximity of C_1 and B_1 , the motion from C_1 to C_2 captures the motion starting from B_1 , and forces this motion to go to B_2 . This is represented by the open arrows. (From the cover principle¹⁴ a motion from A_1 to A_2 is also induced.) We experimented with this paradigm and found that humans experience these illusions²³.

There are many possible forms of smoothing, but two restrictions apply. To ensure that speed decreases with distance (Fig. 1), that is, that the interaction falls off at infinity, c_0 must be non-zero²³. Also, c_m must be non-zero for some $m > 1$, so that solutions may exist for sparse data²⁴. Previous theories that integrated motion over space^{10,19} used only first-order derivatives and did not satisfy these criteria. The higher-order derivatives affect the rate of fall of the interaction at infinity. For this paper's simulations, we chose the smoothing such that the interaction is a gaussian. This form of interaction was chosen for four reasons: (1) It meets the criteria above; (2) it generates analytical solutions; (3) it has a natural spatial scale; and (4) it approximates several processes in vision²⁰. Moreover, variations of the higher-order terms, c_m for $m > 2$, seems to have little effect on the interaction, which thus almost always resembles a gaussian²⁰. This suggests that although analytical solutions may not be

possible for other choices of smoothing, good approximations with lower terms are possible. Thus, these choices are probably not computationally prohibitive. To obtain a gaussian interaction we set $c_m = \sigma^{2m}/(m!2^m)$ (ref. 23). In this case, the solution of equation (1), obtained by standard calculus of variations, is

$$\vec{v}(\vec{r}) = \sum_i \frac{\vec{\beta}_i}{2\pi\sigma^2} \exp\left(-\frac{|\vec{r} - \vec{r}_i|^2}{2\sigma^2}\right) \quad (3)$$

where the $\vec{\beta}_i$ may be computed analytically. This solution is the superposition of gaussians centred in points where motion information is available.

This paper's first example (Fig. 1) shows that the motion coherence theory may account for motion cooperativity^{3,4}. When dots move randomly, but slightly biased, a coherent motion percept in the bias direction occurs. The theory predicts this behaviour, because the smaller variance of the computed velocity field makes the mean motion more perceptible.

In another example, it is shown in Fig. 2 that the motion coherence theory provides a solution to the aperture problem⁸, which can be considered a problem of coherent motion¹⁶. In this particular example, the problem is to find the motion of a contour, because local motion detectors can only detect motion that is perpendicular to the contour. We apply smoothness to the velocity field in two-dimensional space, in contrast to Hildreth¹⁹ who applies smoothness along the contour. (There is evidence that in humans, however, the motion system prefers integration of information along contours²⁵.) The solution has the same form of equation (3), but with an integral along the contour replacing the sum. The new method can provide a good solution to the aperture problem; in the example, the estimated speed is only 0.4% off the true speed. For the gaussian interaction, however, the computed speed is never exactly correct, being always smaller than the true one. This is due to the term c_0 of equation (1), which forces the overall speed down, but is necessary to ensure that the spatial interactions fall off smoothly at infinity²³. (Without this term, it can be proven that the correct solution may be obtained for a certain class of stimuli²³ with similar results holding for Hildreth's method²⁶.) The 'incorrect' computed speed may explain the apparent non-rigidity of certain translating smooth planar curves²⁵.

A final example (Fig. 3) shows a possible explanation for motion capture^{5,6}. In these phenomena, randomly moving dots are captured and move coherently with a superimposed grating or a surrounding contour. The motion coherence theory simultaneously solves the aperture problem of the surrounding contour and captures the internal dots. The theory also predicts that capture may happen when only a few dots move. An example of this occurs when a central ambiguous motion is captured by a peripheral unambiguous motion (Fig. 3c).

In conclusion, a theory for motion coherence that deals simultaneously with the aperture problem and the phenomena of motion cooperativity and motion capture is proposed. The theory differs from other recent works that deal with similar problems^{27,28}. While these works are based on neural considerations, we provide a computational theory expressed as an optimization process. Theories derived in this way can directly incorporate the natural constraints of the visual world²⁹. In spite of the computational approach, the theory's elements can be interpreted in terms of neural processes.

We thank E. C. Hildreth and E. H. Adelson for useful comments on the manuscript. N.M.G. was supported by ONR, and the Sloan and Fairchild foundations. A.L.Y. was supported by USARO.

Received 29 December 1987; accepted 17 March 1988.

1. Ternus, J. *Psychol. Forsch.* **7**, 81-136 (1926).
2. Koffka, K. *Principles of Gestalt Psychology* (Harcourt, Brace & World, New York, 1935).
3. Williams, D. W. & Sekuler, R. *Vision Res.* **24**, 55-62 (1984).
4. Williams, D. W., Phillips, G. & Sekuler, R. *Nature* **324**, 253-255 (1986).
5. Mackay, D. M. *Nature* **192**, 739-740 (1961).

6. Ramachandran, V. S. & Anstis, S. M. *Vision Res.* **23**, 1719-1724 (1983).
7. Verri, A. & Poggio, T. *Artif. Intell. Memo* **917**, (MIT, Cambridge, 1986).
8. Marr, D. & Ullman, S. *Proc. R. Soc. B* **211**, 151-180 (1981).
9. Braddick, O. J. *Phil. Trans. R. Soc. Lond.* **B290**, 137-151 (1980).
10. Horn, B. K. P. & Schunk, B. G. *Artif. Intell.* **17**, 185-203 (1981).
11. Hassenstein, B. & Reichardt, W. E. *Z. Naturforsch.* **11b**, 513-524 (1956).
12. van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am.* **A1**, 451-473 (1984).
13. Adelson, E. H. & Bergen, J. R. *J. opt. Soc. Am.* **A2**, 284-299 (1985).
14. Ullman, S. *The Interpretation of Visual Motion* (MIT Press, Cambridge, 1979).
15. Grzywacz, N. M. & Yuille, A. L. *AIP Conf. Proc.* **151**, (ed. Denker, J. S.) 200-205 (American Institute of Physics, New York, 1986).
16. Adelson, E. H. & Movshon, J. A. *Nature* **300**, 523-525 (1982).
17. Anstis, S. M. *Vision Res.* **10**, 1411-1430 (1970).
18. Ramachandran, V. S. & Anstis, S. M. *Vision Res.* **23**, 83-85 (1983).
19. Hildreth, E. C. *The Measurement of Visual Motion* (MIT Press, Cambridge, 1983).
20. Poggio, T., Voorhes, H. & Yuille, A. L. *Artif. Intell. Memo* **776**, (MIT, Cambridge, 1984).
21. Bertero, M., Poggio, T. & Torre, V. *Artif. Intell. Memo* **924**, (MIT, Cambridge, 1987).
22. Ramachandran, V. S. & Inada, V. *Spatial Vision* **1**, 57-67 (1985).
23. Yuille, A. L. & Grzywacz, N. M. *Artif. Intell. Memo* **1002**, (MIT, Cambridge, 1987).
24. Duchon, J. *Lecture Notes in Math.* **571**, (eds Schempp, W. & Zeller, K.) 85-100 (Springer-Verlag, Berlin, 1979).
25. Nakayama, K. & Silverman, G. H. *Invest. Ophthalmol. Vis. Sci.* **26**, 188 (1985).
26. Yuille, A. L. *Artif. Intell. Memo* **724**, (MIT, Cambridge, 1983).
27. Little, J., Bulthoff, H. & Poggio, T. *Proc. Image Understanding Workshop* (ed. Baumann, L.) 915-920 (Science Applications International Corporation, McLean, Virginia, 1987).
28. Heeger, D. J. *J. opt. Soc. Am.* **A4**, 1455-1471 (1987).
29. Marr, D. *Vision* (Freeman, New York, 1982).
30. von Schiler, P. *Psych. Forsch.* **17**, 179-214 (1933).

A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains

David C. Kaslow^{*,†}, Isabella A. Quakyi^{*}, Chiang Syin^{*}, Michael G. Raum[†], David B. Keister^{*}, John E. Coligan[†], Thomas F. McCutchan^{*} & Louis H. Miller^{*}

^{*} The Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, Maryland 20892, USA

[†] The Biological Resources Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

[‡] The Interinstitute Medical Genetics Program, Clinical Center, National Institutes of Health, Bethesda, Maryland 20892, USA

Malaria vaccines are being developed against different stages in the parasite's life cycle¹, each increasing the opportunity to control malaria in its diverse settings. Sporozoite vaccines are designed to prevent mosquito-induced infection²; first generation recombinant or synthetic peptide vaccines have been tested in humans^{3,4}. Asexual erythrocytic stage vaccines, developed to prevent or reduce the severity of disease, have been tested in animals^{1,5} and in humans⁶. A third strategy is to produce sexual stage vaccines that would induce antibodies which would prevent infection of mosquitoes when ingested in a bloodmeal containing sexual stage parasites¹. Although not directly protective, the sexual stage vaccine combined with a sporozoite or asexual stage vaccine (protective component) could prolong the useful life of the protective component by reducing transmission of resistant vaccine-induced mutants. In areas of low endemicity, the sexual stage vaccine could reduce transmission below the critical threshold required to maintain the infected population, thereby assisting in the control or eradication of malaria. Transmission of *Plasmodium falciparum*, the major human malaria, can be blocked by monoclonal antibodies against three sexual stage-specific antigens⁷⁻⁹. We have cloned the gene encoding the surface protein of relative molecular mass *M*_r 25,000 (25K; Pfs25), expressed on zygotes and ookinetes of *P. falciparum*. The deduced amino-acid sequence consists of a signal sequence, a hydrophobic C-terminus, and four tandem epidermal growth factor EGF-like domains¹⁰⁻¹².

Difficulties have been encountered in cloning the genes encoding the three sexual stage-specific antigens of *P. falciparum*,

partly because the mAb-defined epitopes are dependent on disulphide bonds⁷ and partly because of problems in obtaining the large quantities of parasites and purified protein necessary for peptide sequencing. Our cloning strategy was to purify Pfs25 from zygotes of *P. falciparum* (3D7 clone of NF54) using immunoaffinity chromatography and SDS-PAGE, and then to microsequence the protein so that we could synthesize oligonucleotide probes to screen genomic DNA libraries. Because the N-terminus was blocked, purified Pfs25 was digested with trypsin, and the resulting peptides were fractionated by HPLC. Five peptide sequences were obtained (Fig. 1).

From a tryptic peptide sequence (Fig. 1, double lines), a highly degenerate oligonucleotide probe was synthesized and used to clone a 600 base pair (bp) *Dra*I fragment (pSKR 2) of genomic DNA. The *Dra*I fragment contained one long open reading frame, but no termination codon. Therefore a 3.5 kilobase (kb) *Hind*III fragment (pNF4.13) was cloned. Both cloned fragments were used to determine the nucleotide sequence of Pfs25.

In Fig. 1, the deduced amino-acid sequence for Pfs25 is shown above the nucleotide sequence. The gene has a single exon that codes for a polypeptide of 217 amino acids. Proof that we have cloned the gene for the 25K ookinete surface antigen is as follows. First and most important, all the tryptic peptides of Pfs25 that we microsequenced are found within the deduced amino-acid sequence of Pfs25 (Fig. 1). Furthermore, radioactive peaks for Cys and Met in the sequence of a metabolically labelled tryptic peptide of Pfs25 perfectly matched the position of these residues in the deduced amino-acid sequence of the Pfs25 gene (see asterisks, Fig. 1). Second, the gene is expressed preferentially in the sexual stages of *P. falciparum*, the stage in which Pfs25 is synthesized⁷. Messenger RNA from gametocytes (not shown) and 5-h-old zygotes (Fig. 2, lane 3) of *P. falciparum* (3D7 clone of NF 54; ref. 13) have an abundance of an mRNA species (~1.4kb) that hybridize to pSKR 2, whereas mRNA from asexual stage parasites (3D7; Fig. 2, lane 2) gives only a weak signal. The weak signal from the asexual stage parasites is probably due to the presence of some gametocytes in the preparation, as pSKR 2 does not hybridize with mRNA (Fig. 2, lane 1) from a *P. falciparum* parasite that produces no gametocytes (the LF4 clone of a Liberian isolate of *P. falciparum*¹³). Third, when gametes, zygotes or ookinetes of *P. falciparum* are metabolically labelled, Pfs25 is the predominant product incorporating [³⁵S] cysteine (N. Kumar and R. Carter, unpublished results) but is poorly labelled with [³⁵S] methionine. In correlation with this, the deduced amino-acid sequence contains 11% cysteine. Fourth, the structure of the protein from the deduced amino-acid sequence is consistent with previous reports about the biochemistry of the 25K surface glycoprotein¹⁵. The deduced sequence contains a putative signal peptide at the N-terminus, and a short hydrophobic anchor at the C-terminus; it has four potential glycosylation sites for N-linked sugars; and it encodes a polypeptide of predicted *M*_r 24.1K. The short hydrophobic anchor (15 amino acids), the lack of a potential cytoplasmic (hydrophilic) region at the C-terminus, and the observation that Pfs25 contains a covalently attached fatty acid¹⁵, suggest that Pfs25 may have a phosphatidyl-inositol-like membrane anchor¹⁶.

Between the signal sequence and the hydrophobic C-terminus, the organization of the cysteines is indicative of four tandemly repeated EGF-like domains (Fig. 3a). As in human EGF-precursor¹⁰⁻¹² and human LDL-receptor¹⁷, three cysteine residues precede the consensus sequence Y/F-x-C-x-C-x-x-G-Y/F, and one follows it (Fig. 3b). The EGF-like domain is also found in invertebrates (for example the notch protein¹⁸ in *Drosophila melanogaster* and lin-12 in *Caenorhabditis elegans*¹⁹). This is the first report of the presence of EGF-like domains in proteins of unicellular organisms. A vaccinia virus growth factor which shares amino-acid sequence homology with EGF induces proliferation of unaffected host cells surrounding plaques²⁰ and increases virulence²⁷. The EGF-like domains of

Fig. 1 Nucleotide and predicted amino-acid sequence of Pfs25. The predicted coding region of the nucleotide sequence is shown in capitals and consists of 654 bp. Because the N-terminus is blocked, the start of the mature protein has not been determined. Solid lines, tryptic peptide sequences determined by microsequencing; asterisks, sequence determined by microsequencing radiolabelled peptides; dotted lined, indeterminate amino-acid residue of microsequenced peptide; double solid line, tryptic peptide sequence used to construct oligonucleotide probes; open circles, sites of possible asparagine-linked glycosylation; broken lines, hydrophobic regions.

Methods. 3D7 clone of NF 54 isolate of *P. falciparum* was cultured *in vitro* and zygotes were prepared as described⁹. Using mAb antibody 1C7 covalently linked to Sepharose 4B beads (CNBr-activated), Pfs25 was immunoadfinity purified from Triton X-100 extracts of 5-h-old zygotes (10⁹) which had been metabolically labelled with Trans S35 (70% [³⁵S] methionine, 20% [³⁵S] cysteine, and 10% other [³⁵S] compounds; ICN Radiochemicals). The beads containing Pfs25 were resuspended in SDS-PAGE sample buffer (5% SDS, 62.5 mM Tris, pH 6.8, and 0.002% Bromophenol blue) containing 8 M urea, and were heated at 68 °C for 5 min. The eluted protein in sample buffer was loaded on to a 12% SDS-polyacrylamide gel under non-reducing conditions. Pfs25 was the only radiolabelled protein identified on the gel and was recovered from the gel by passive diffusion. The lyophilized sample was resolubilized in 0.5 M Tris-HCl, pH 8.5, 6 M guanidine hydrochloride, 0.3 mM EDTA buffer containing 64 mM dithiothreitol, and incubated for 2 h at 37 °C under nitrogen. Iodoacetamide was added to a final concentration of 174 mM and samples were left for 1 h at 25 °C in the dark. An excess of 2-mercaptoethanol was added, followed by 10 vols of absolute ethanol. The reduced and alkylated protein was allowed to precipitate at -20 °C for 4 h. The remaining pellet was resuspended in 50 mM NH₄HCO₃, pH 7.9 and digested with two 0.5 µg doses of TPCK-treated trypsin (Sigma), for 6 h at 37 °C before heating to 65 °C for 10 min. The tryptic peptides were fractionated on a reverse phase HPLC (RP-300, Applied Biosystems) using an 0-50% acetonitrile gradient with 0.1% trifluoroacetic acid. Peptide microsequencing was performed on 1% of each cycle's product was analysed on an attached model 120A peptide, and the radioactivity of the remaining 60% of each cycle's product 1, 7, 12 and 17. The peptide sequence identified by double solid line, varying the codon for proline, the digonucleotides were divided into four bands of zygote RNA and to a 600-bp *Dra*I fragment of genomic DNA screened and one clone identified and subcloned into Bluescript SK (phage size-selected library in the vector pSP64. Both strands of each clone were

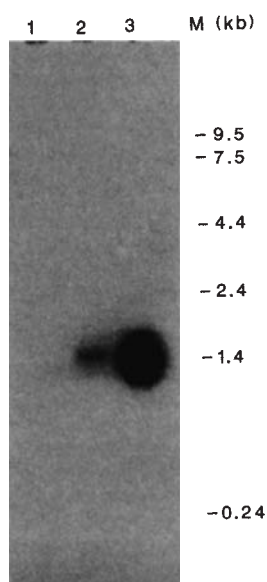


Fig. 2 Northern blot analysis of asexual and sexual stage RNA. Total cellular RNA (20 µg per lane) was prepared²⁴ from asexual parasites of LF4 (lane 1), asexual parasites of 3D7 (lane 2), or 5-h-old zygotes of 3D7 (lane 3). Samples were electrophoresed through a 1% agarose/formaldehyde gel and transferred to a nylon membrane. The filter was hybridized at 55 °C overnight with randomly primed pSKR2 insert (specific activity 10⁹ c.p.m. µg⁻¹), and washed as described²⁵. M, RNA size markers in kb (BRL).

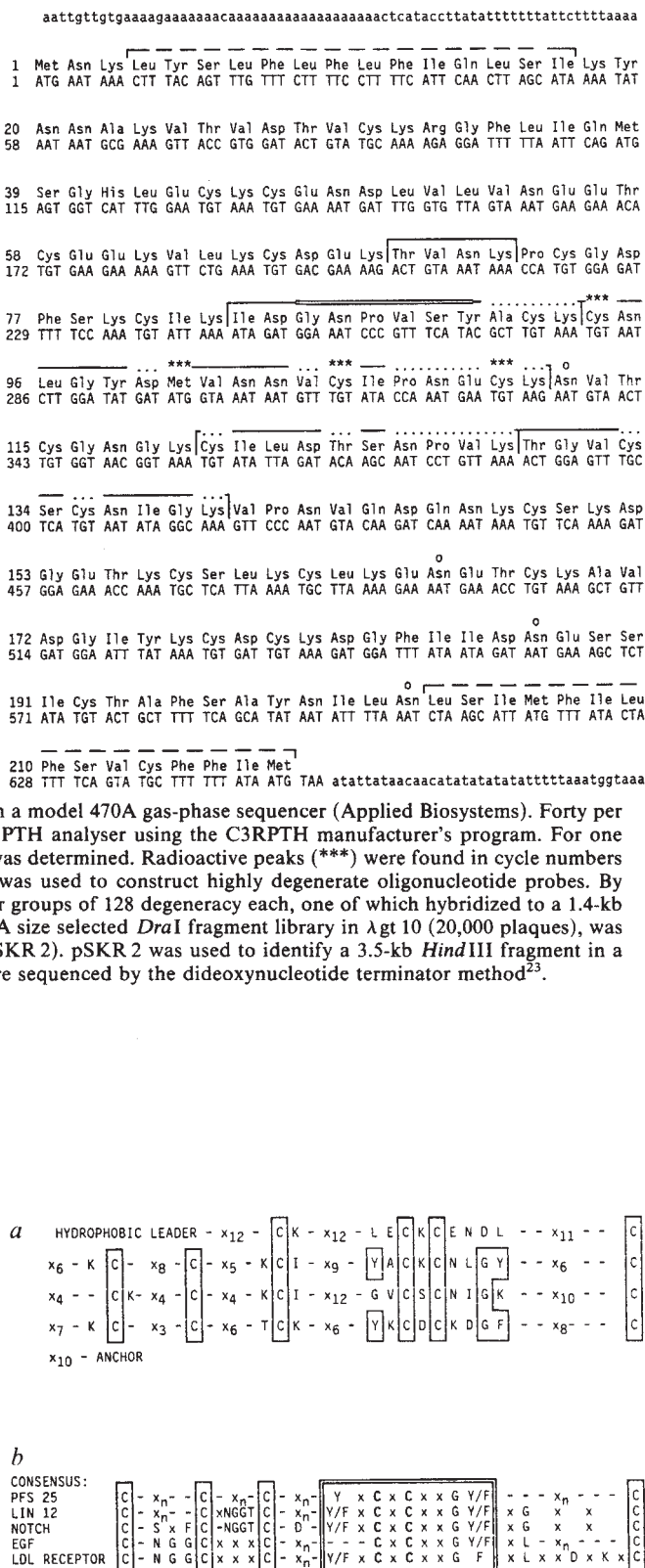


Fig. 3 *a*, Protein structure of Pfs25 arranged to emphasize the relatedness of the four EGF-like domains. Cysteine residues, and other identical or related amino acids are boxed. *b*, EGF-like repeat consensus sequence from Pfs25, lin-12¹⁹, notch¹⁸, EGF^{10,11} and human LDL-receptor¹⁷. Cysteine residues are boxed. A double box frames the core of the consensus sequence C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; I, Ile; K; Lys; L, Leu; N, Asn; T, Thr; Y, Tyr: x, any amino acid; ---, unspecified number of amino acids.

malaria ookinetes may bind to receptors on mosquito epithelial cells to facilitate invasion. Of more practical concern is the use of these shared amino-acid sequences in synthetic peptide or recombinant protein vaccines, as shared sequences between host and pathogen have been implicated in certain severe autoimmune complications^{20,21}. During preliminary vaccine trials of Pfs25, recipients should be monitored for development of auto-antibodies.

Malaria continues to exact a heavy toll. In rural Africa, approximately 25% of all deaths in children 1–4 yr of age are caused by malaria²², despite the sensitivity of local parasites to chloroquine. Mosquito control is difficult in this setting. At present, the greatest hope for reducing this mortality is a protective vaccine that reduces disease and death by suppressing the replication of the parasite. Vaccines designed to induce transmission-blocking immunity to Pfs25 and other sexual stage antigens would prolong the usefulness of such protective vaccines, as well as reduce the spread of malaria in areas of low transmission.

We thank R. Carter for critical discussion. I.Q. and C.S. are recipients of financial support from UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

Received 4 February; accepted 13 March, 1988.

1. Miller, L. H. *et al. Science* **234**, 1349–1356 (1986).
2. Zavala, F. *et al. J. exp. Med.* **166**, 1591–1596 (1987).
3. Ballou, W. R. *et al. Lancet* **8545**, 1277–1281 (1987).
4. Herrington, D. A. *et al. Nature* **328**, 257–259 (1987).
5. Patarroyo, M. E. *et al. Nature* **328**, 629–632 (1987).
6. Patarroyo, M. E. *et al. Nature* **332**, 158–161 (1988).
7. Rener, J., Graves, P. M., Carter, R., Williams, J. L. & Burkot, J. R. *J. exp. Med.* **158**, 976–981 (1983).
8. Vermeulen, A. N. *et al. J. exp. Med.* **162**, 1460–1476 (1985).
9. Quakyi, I. *et al. J. Immunol.* **139**, 4213–4217 (1987).
10. Gray, A., Dull, T. J. & Ullrich, A. *Nature* **303**, 722–725 (1983).
11. Scott, J. *et al. Science* **221**, 236–240 (1983).
12. Doolittle, R. F., Feng, D. F. & Johnson, M. S. *Nature* **307**, 558–560 (1984).
13. Walliker, D. *et al. Science* **231**, 1661–1666 (1987).
14. Graves, P. M., Carter, R., Keystone, J. F. & Seely Jr, D. C. *Am. J. trop. Med. Hyg.* **33**, 212–219 (1984).
15. Vermuelen, A. N. *et al. Molec. biochem. Parasitol.* **20**, 155–163 (1986).
16. Low, M. G. & Saltiel, A. R. *Science* **239**, 268–275 (1988).
17. Sudhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. *Science* **228**, 815–822 (1985).
18. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. *Cell* **43**, 567–581 (1985).
19. Greenwald, I. *Cell* **43**, 583–590, (1985).
20. Jahnke, U., Fisher, E. H. & Alvord, E. C. *Science* **229**, 282–284 (1985).
21. Fujinami, R. S. & Oldstone, M. B. A. *Science* **230**, 1043–1045 (1985).
22. Greenwood, B. M. *et al. Trans. R. Soc. trop. Med. Hyg.* **81**, 478–486 (1987).
23. Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165–170 (1985).
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* 194–195 (Cold Spring Harbor Laboratory, New York, 1982).
25. Kaslow, D. C. *et al. Genomics* **1**, 19–28 (1987).
26. Buller, R. M. L., Chakrabarti, S., Moss, B. & Fredrikson, T. *Virology* (in the press).
27. Buller, R. M. L., Chakrabarti, S., Cooper, J. A., Ewardczik, D. R. & Moss, B. *J. Virol.* **62**, 866–874 (1988).

Paramyosin and actin in schistosomal teguments

Yoshitsugu Matsumoto*, George Perry*,
Rhea J. C. Levine†, Ronald Blanton‡,
Adel A. F. Mahmoud‡ & Masamichi Aikawa*§

* Institute of Pathology and † Department of Medicine,
Case Western Reserve University, Cleveland, Ohio 44106, USA

‡ Department of Anatomy, The Medical College of Pennsylvania,
Philadelphia, Pennsylvania 19129, USA

Schistosomes are blood-dwelling trematode parasites that infect 200 million people in developing countries. The critical role served by the tegument in immune evasion and parasite homeostasis suggests that a detailed knowledge of tegumental components would be helpful in the design of new drugs and the production of vaccines. We demonstrate here, by immunoelectron microscopy, that the cytoskeletal proteins actin and paramyosin are organized into major tegumental structures of *Schistosoma mansoni*. The

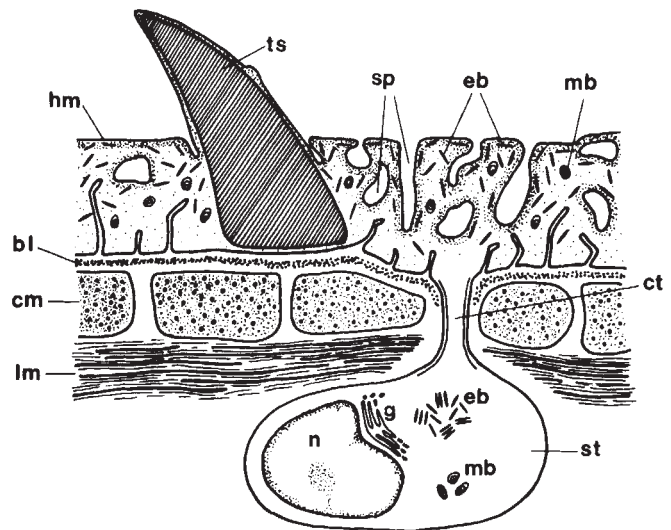


Fig. 1 Diagrammatic representation of the adult schistosome tegument and underlying structures. bl, basal lamina; cm, circularly arranged muscle fibres; ct, cytoplasmic tube; eb, elongate body; g, Golgi complex; hm, heptalaminate outer membrane; lm, longitudinal muscle fibres; mb, membranous body; n, nucleus of subtegumental cell; sp, surface pits; st, subtegumental cell (cyton); ts, tegumental spine.

surface spines are composed of paracrystalline arrays of actin filaments. Actin is also present in areas recovering from damage, implying an important role for this structural protein in tegumental repair. Paramyosin exists predominantly in the tegument in a non-filamentous form, the membrane-bounded elongate bodies. The localization of this protein to the tegument of the parasite is the likely basis for resistance to *S. mansoni* observed in mice immunized with paramyosin (refs 1, 2 and T. P. Flanigen *et al.*, in preparation).

The extensive recent efforts devoted to characterization of the proteins and genes encoding vital schistosome components^{3,4} are directed toward production of a vaccine capable of inducing significant immunity to the parasite. The finding that mice injected with extracts of *S. mansoni* produce antibodies to the invertebrate muscle protein paramyosin^{1,2} led us to examine the localization of this protein and another muscle protein actin.

From free-swimming cercaria through the adult worm stage, the tegument of *S. mansoni* consists of a cytoplasmic syncytium ~2–4 µm thick^{5–7}, connected by cytoplasmic processes to nucleated cell bodies (cytons) located beneath the cortical muscle (Fig. 1). Membranous bodies and elongate bodies are found scattered in the matrix of the tegument and also in association with the Golgi apparatus in the cytons. Tegumental spines are another conspicuous feature of the schistosome surface. In negatively stained specimens, electron microscopy reveals a 38.5 nm axial periodicity, similar to that of actin bundles, along filament bundles found in the spinal cores⁸. At the light microscopic level, actin has been demonstrated in the spines by antibody and chemical staining^{9,10}.

We have used an affinity purified rabbit antiserum (AGA) and a mouse monoclonal antibody to chicken gizzard actin (C4) (both of which recognize all actin isoforms¹¹), coupled with indirect colloidal gold labelling, to demonstrate at the ultra-microscopic level the presence of actin in both spines and muscle. Actin immunoreactivity in the spines is clearly limited to the electron-dense paracrystalline cores (Fig. 2a). An identical pattern of anti-actin labelling is associated with the spines of cercariae, schistosomula and the adult female (data not shown). Cortical muscle located immediately beneath the tegument is also intensely labelled. Actin is not detected in the tegument matrix except for nerve ending bulbs (ref. 7 and data not shown) and in regions (Fig. 2b) where tegumental damage

§ To whom correspondence should be addressed.

Fig. 2 Immunoelectron micrographs showing localization of actin in tegumental spine (a) and damaged region of tegument on the dorsal surface of a *S. mansoni* male adult (b). m, muscle; scale bars, 0.5 μm .

Methods. Male adult worms of a Puerto Rican strain of *S. mansoni* were obtained by portal vein perfusion of 8-week infected Cf1 mice. Dulbecco's minimal essential medium induced tegumental damage¹². Parasites were fixed with 1% paraformaldehyde and 0.1% glutaraldehyde in 0.1 M sodium phosphate buffer (PB) at pH 7.3 for 30 min at room temperature. Samples were then washed in 0.1 M PB, dehydrated in a graded series of ethanols, embedded in LR gold (London Resin Co.) and polymerized at -20°C for 30 h under ultraviolet light. Thin sections (70–90 nm) were cut with a diamond knife and mounted on nickel grids. Sections were etched with a saturated aqueous solution of sodium metaperiodate for 1 h at 37°C , then washed and incubated with 0.1 M PB containing 0.9% NaCl (PBS), 5% nonfat dry milk and 0.01% Tween 20 (PBS-Milk-Tween) for 30 min at room temperature. Grids were then transferred to a solution of mouse monoclonal antibody to chicken gizzard actin (C4, ref. 11) diluted at $50\text{ }\mu\text{g ml}^{-1}$ in PBS-Milk-Tween and incubated for 1 h. Sections and schistosome lysate immunoblots showed no reactivity to chicken gizzard actin-adsorbed C4 antibody. Control sections were incubated with the same dilution of an unrelated mouse monoclonal antibody (not shown). The grids were rinsed in PBS containing 1% BSA and 0.01% Tween 20 (PBS-BSA-Tween) and transferred to a drop of goat anti-mouse IgG conjugated with gold particles (15 nm, Janssen) diluted at 1 in 20 with PBS-Milk-Tween. After 1 h of incubation, the grids were rinsed in PBS-BSA-Tween followed by distilled water. The sections were dried and stained with 2% uranyl acetate in 50% methanol and examined by a JEOL 100 CX electron microscope.

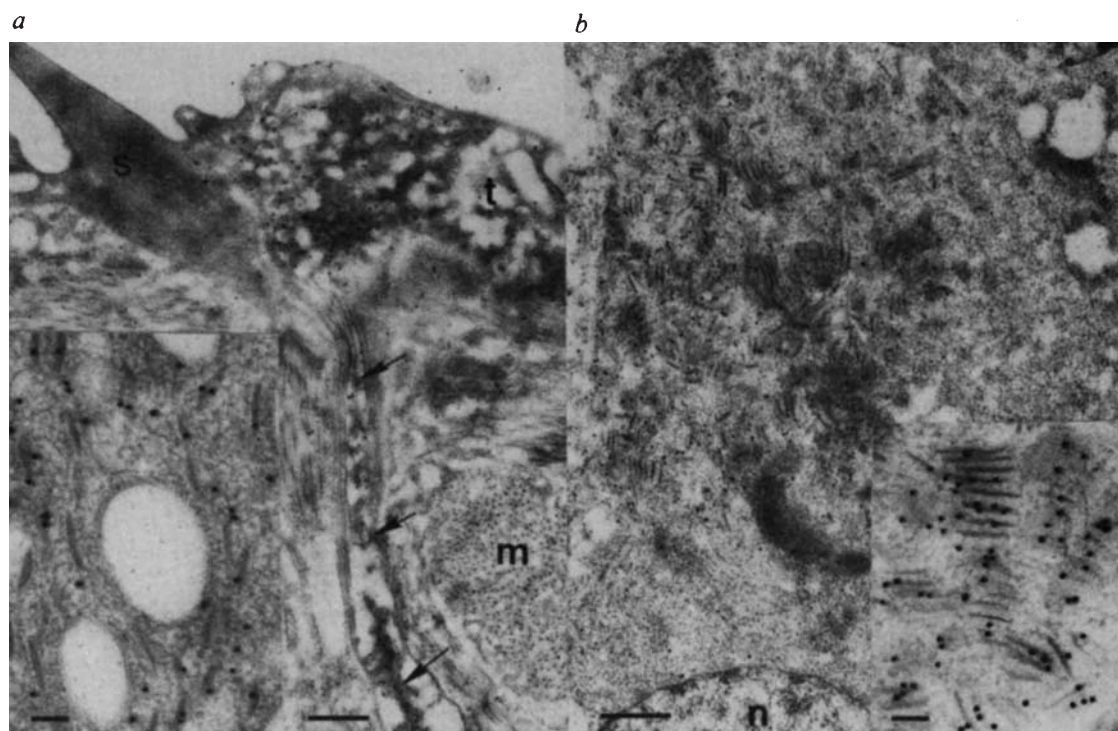
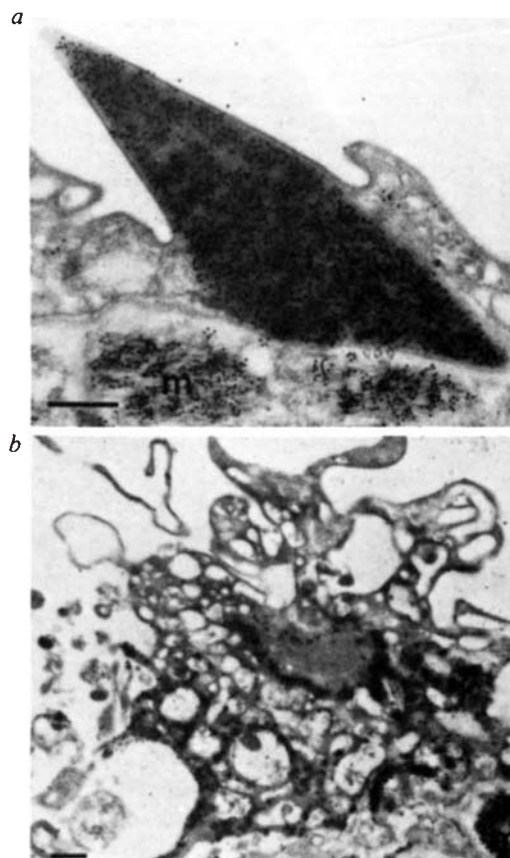


Fig. 3 Immunoelectron micrographs demonstrating paramyosin in the tegument and cyton of male adult *Schistosoma mansoni*. Gold labelling is in the tegument (a), cyton (b) and cytoplasmic tube (arrows in a) by which the tegument is connected to cytons. Most immunoreactivity is observed in association with the elongate inclusions of the tegument (inset of a) and cyton (inset of b). m, muscle; n, nucleus of subtegumental cell; s, spine; t, tegument. Scale bars, 0.5 μm . Scale bars in insets, 0.1 μm .

Methods. Sections were incubated with rabbit affinity-purified IgG purified IgG to *Limulus* (horseshoe crab) muscle paramyosin¹⁷ diluted at 1 in 320 with PBS-Milk-Tween. The antibody was adsorbed with both native *Limulus* tropomyosin and native *Limulus* actomyosin. Control sections or immunoblots of schistosome lysate incubated with the same dilutions of the unrelated rabbit serum or *Limulus* adsorbed paramyosin antibody ($200\text{ }\mu\text{g ml}^{-1}$) showed no reactivity (not shown). After washing in PBS-BSA-Tween, the sections were incubated with goat anti-rabbit IgG conjugated with gold particles, rinsed, dried and stained as described in Fig. 2.

and healing^{12,13,15} is morphologically apparent. Figure 2b shows a section of adult male tegument with a discontinuity bordered by a region of increased vacuolarization, microvillar projections and condensation of the cytoplasmic processes between the vacuoles. There is a striking accumulation of actin within these dense homogenous areas of the tegumental matrix.

Tegument repair processes are of vital importance in maintaining the integrity of the schistosome tegument which is vulnerable to damage by mechanical or pharmacological agents and by the host's immune system^{14,16}. Trematodes have active tegumental repair mechanisms^{12,14} that demonstrate little dependence on new protein synthesis¹³. The distribution of actin we observe and reports of resorption of tegumental spines in response to immunologic, environmental, or chemical stresses^{14,16} suggest that repair involves the use of actin and possibly other structural proteins released from tegument spines.

Whereas actin and its associated proteins localize intensely to filamentous structures in both tegument and muscle, the IgG fraction of a rabbit polyclonal antiserum specific for *Limulus* paramyosin¹⁷ rarely binds to the thick filaments of cortical muscle. Instead, these antibodies bind almost exclusively to the contents of the non-filamentous membrane-bounded elongate bodies within the tegument and the cytons (Figs. 3a and 3b). Rabbit polyclonal antisera to *Owenia* paramyosin¹⁸ give identical results. Binding of paramyosin antibodies to the teguments of cercariae and schistosomulae is less than that observed in adults, a result consistent with the lesser number of elongate bodies in these stages. Furthermore, prominent gold labelling is observed in the esophageal lining which is continuous with the tegumental syncytium and which contains elongate bodies similar to those in the tegument. Immunoblots of *S. mansoni* proteins and purified *Limulus* paramyosin¹⁷ stained with the same antibodies demonstrate that paramyosin is also a major component of the parasite (data not shown). Antibodies previously adsorbed with purified *Limulus* paramyosin fail to stain schistosomal structures or to recognize paramyosin on immunoblots, confirming the specificity of this immune reaction (data not shown).

The localization of paramyosin in a non-filamentous membrane-bounded form in non-muscle tissue is a novel and quite unexpected result. Paramyosin is usually organized into the core structure of the thick filaments of invertebrate muscle, where it probably forms a scaffold into which cortical myosin molecules attach¹⁹⁻²¹. It may also be involved in the regulation of catch contractions in molluscan smooth muscles²². The presence of this protein in the elongate bodies of the schistosome tegument and cytons suggests that it may be a secretory product of this tissue, and further, that its distribution and function(s) are more diverse than is currently known. The abundance of paramyosin in the outermost layer of the parasite may explain the protective immunity to schistosome infection induced in mice immunized with this protein (refs 1, 2 and Flanigen *et al.*, in preparation). Our results are of direct relevance to the development of an anti-schistosomal vaccine or chemotherapy and suggest that approaches directed against paramyosin, which is absent from vertebrate host organisms, may prove efficacious in controlling schistosomiasis.

We thank James Lessard for the antibodies to actin, Monique Diano for the antisera to *Owenia* paramyosin, and Paul Mulvihill and Kiet Dan Luc for assistance in these studies. Supported by grants from the US Public Health Service and UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

Received 15 February; accepted 29 March 1988.

1. Pearce, E. J., James, S. L., Dalton, J., Barrall, A., Ramas, C., Strand, M. & Sher, A. *J. Immunol.* **137**, 3593-3600 (1986).
2. Lanar, D. E., Pearce, E. J., James, S. L. & Sher, A. *Science* **234**, 593-596 (1986).
3. Capron, A., Dessaint, J. P., Capron, M., Ouma, J. H. & Butterworth, A. E. *Science* **238**, 1065-1072 (1987).
4. Simpson, A. J. G. & Cioli, D. *Parasit. Today* **3**, 26-28 (1987).

5. Morris, G. P. & Threadgold, L. T. *J. Parasit.* **54**, 15-27 (1968).
6. Smith, J. H., Reynolds, E. S. & von Lichtenberg, F. *Am. J. trop. Med. Hyg.* **18**, 28-49 (1969).
7. Hockley, D. J. & McLaren, D. J. *Int. J. Parasit.* **3**, 13-25 (1973).
8. Cohen, C., Reinhardt, B., Castellani, L., Norton, P. & Stirewalt, M. *J. Cell Biol.* **95**, 987-988 (1982).
9. Davis, A. H., Blanton, R. & Klich, P. *Molec. Biochem. Parasit.* **17**, 289-298 (1985).
10. Abbas, M. K. & Cain, G. D. *Parasit. Res.* **73**, 66-74 (1987).
11. Galloway, P. G., Perry, G. & Gambetti, P. *J. Neuropathol. exp. Neurol.* **46**, 185-199 (1987).
12. Carlisle, S., Weisberg, L. S. & Bentley, A. G. *J. Parasit.* **69**, 319-334 (1983).
13. Bentley, A. G. *J. Parasit.* **68**, 1096-1104 (1982).
14. Shaw, M. K. & Erasmus, D. A. *Parasitology* **94**, 243-254 (1987).
15. Kassis, A. I., Aikawa, M. & Mahmoud, A. A. F. *J. Immun.* **122**, 398-405 (1979).
16. McCormick, S. L. & Damian, R. T. *J. Parasit.* **73**, 130-143 (1987).
17. Elfvin, M., Levine, R. J. C. & Dewey, M. M. *J. Cell Biol.* **71**, 261-272 (1976).
18. Diano, M., Bivic, A. L. & Hirn, M. *Analyt. Biochem.* **166**, 224-229 (1987).
19. Epstein, H. F., Miller, D. M. III, Ortiz, I. & Berliner, G. C. *J. Cell Biol.* **100**, 904-915 (1985).
20. Elliott, A. & Bennett, P. in *Basic Biology of Muscles: A Comparative Approach* (eds Twarog, B., Levine, R. & Dewey, M.) 11-28 (Raven, New York, 1982).
21. Szent-Györgyi, A. G., Cohen, C. & Kendrick-Jones, J. *J. molec. Biol.* **56**, 239-258 (1971).
22. Achazi, R. in *Basic Biology of Muscles: A Comparative Approach* (eds Twarog, B., Levine, R. & Dewey, M.) 291-308 (Raven, New York, 1982).

Activation of human immunodeficiency virus type 1 by DNA damage in human cells

Kristoffer Valerie, Anne Delers*, Claudine Bruck*, Clotilde Thiriart*, Hagai Rosenberg†, Christine Debouck & Martin Rosenberg

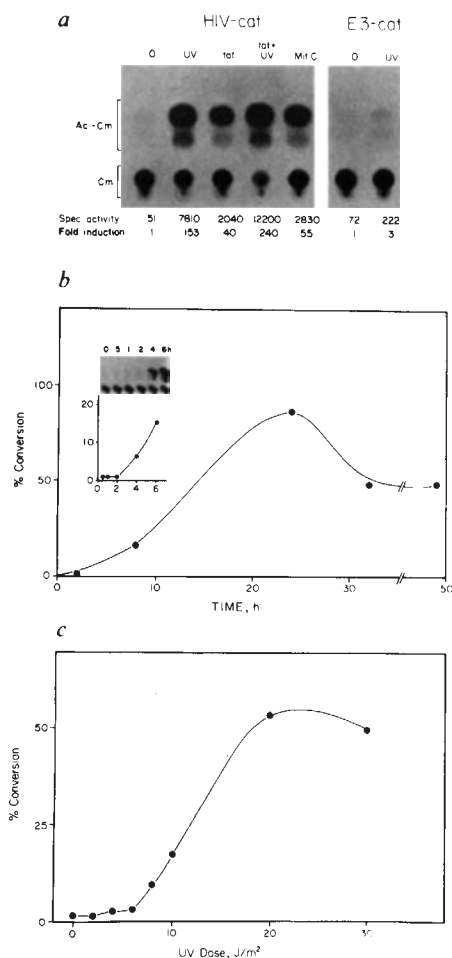
Department of Molecular Genetics, Smith Kline and French Laboratories, Research and Development Division, PO Box 1539, King of Prussia, Pennsylvania 19406-0939, USA

* Department of Molecular and Cellular Biology, Smith Kline-RIT, B-1330 Rixensart, Belgium

Recent studies indicate that human immunodeficiency virus type 1 (HIV) gene expression can be dramatically enhanced by certain heterologous viral^{1,2} and chemical²⁻⁶ agents, implicating these as potential reactivating agents of latent virus infection. A common denominator shared by these agents is their ability to cause stress responses in cells⁷⁻⁹. In an effort to determine whether stress responses affect HIV gene expression, we examined the effects of ultraviolet light (UV) and mitomycin C, on HIV gene expression as well as on viral growth and development. We demonstrate that these agents enhance HIV gene expression up to 150-fold. These levels are similar to those obtained by the *tat* gene product, the HIV trans-activating factor responsible for enhancing viral gene expression¹⁰⁻¹⁵. The increase in gene expression after UV irradiation appears to require transcription but not *de novo* protein synthesis, and correlates with an accumulation of stable mRNA. Most importantly, UV irradiation of human T-cells prior to viral infection significantly shortens the viral growth cycle. Apparently, UV-induced cellular stress is highly conducive for viral replication and growth. We further demonstrate that even direct sunlight can activate HIV gene expression. These results demonstrate that DNA damaging agents, and perhaps other agents which elicit SOS-like stress responses in mammalian cells, can activate HIV expression thereby enhancing viral replication and development.

To examine the effects of DNA-damaging agents on HIV long terminal repeat (LTR)-directed gene expression, a HeLa cell line (HIV-cat) was established which carries an integrated DNA cartridge containing the bacterial *cat* gene under control of the HIV LTR. The HIV-cat HeLa cells were irradiated with UV light and then assayed for chloramphenicol acetyltransferase (CAT) activity. Maximal activation of *cat* gene expression ranged from 50- to 150-fold in various experiments (Fig. 1). Analogous experiments carried out using mitomycin C gave induction levels of about 50-fold (Fig. 1a). In order to compare

† Permanent address: Israel Institute for Biological Research, PO Box 19, Ness-Ziona 70450, Israel.



these levels of activation with those induced by the *tat* protein, we transfected HIV-cat cells with a plasmid carrying the HIV *tat* gene under the transcriptional control of the Rous sarcoma virus LTR promoter region and monitored expression of CAT activity. We reproducibly obtained an increase in *cat* gene expression of 40-fold or more (Fig. 1a), consistent with typically observed levels of *tat*-mediated *trans*-activation¹⁰⁻¹⁵. In contrast, a control HeLa cell line which carried a stably integrated transcription unit containing the adenovirus E3 early promoter fused to the *cat* gene showed little, if any, increase in CAT activity after either UV irradiation (Fig. 1a), or transfection with the *tat* gene plasmid (data not shown). Clearly, both UV and mitomycin C treatment dramatically activate HIV LTR expression, and have little or no effect on E3-promoted expression. Similar results were obtained with three other independent HeLa clones carrying the HIV LTR-*cat* cartridge, indicating that the site of chromosomal insertion was not related to this phenomenon (data not shown). Moreover, when HIV-cat HeLa cells transfected with the *tat* gene were also UV-irradiated, only slightly higher levels of stimulation were observed (~2-fold over UV alone, Fig. 1a). Although it appears that *tat* expression in combination with UV resulted in a slightly synergistic effect on the efficiency of HIV LTR-directed expression, caution must be exercised in evaluating the meaning of this result. It is reasonable to believe that activation of the HIV LTR by UV irradiation affects a larger number of cells than the activation obtained by transfection of *tat* DNA, which is presumably

Fig. 1 a, Activation of the HIV LTR by UV, mitomycin C, and *tat*. HIV-cat and E3-cat HeLa cells were treated as described below. 0, control; UV, UV-irradiated; *tat*, transfected with *tat* DNA; *tat* + UV, transfected with *tat* DNA and UV-irradiated; Mit. C, mitomycin C-treated. Cm, chloramphenicol; Ac-Cm, acetylated chloramphenicol. CAT activities (Spec. activity) are expressed as pmoles of [¹⁴C]-chloramphenicol converted to acetylated forms per min and mg of total protein. b, Time course of UV-induced CAT activity. c, Dose-response of UV-induced CAT activity. HIV-cat HeLa cells were UV irradiated and incubated in medium for the indicated times (b), or irradiated with various doses of UV light (c), and then processed to determine CAT activities (measured as the % conversion of [¹⁴C]-chloramphenicol to its acetylated forms). b, shows the results of two separate experiments.

Methods. To generate stable clones carrying an HIV LTR-*cat* transcriptional unit, HeLa cells were transfected with a 1:5 ratio of pSV2neo²⁰ and pBI120-HIVcatBGH (containing a partial *Bgl*II-*Bam*HI fragment from pHIVcatBGH, which comprises the 565 base pairs (bp) *Bgl*II-*Hind*III fragment (-485 to +80) spanning the 3' HIV LTR (isolated from plasmid p15CAT²¹) upstream of the *cat* gene from pRSVcat²² and bovine growth hormone poly(A) signals²³, cloned in the *Bam*HI site of the plasmid pBI120 [Promega]). Transformed cells were selected for G418 resistance²⁰, expanded and tested for their capacity to be activated by the HIV *tat* gene product by DNA transfection followed by CAT assays. One inducible clone (HIV-cat), was chosen for further studies. Semi-confluent 60-mm dishes, which had been seeded the day before, were treated with 10 J m⁻² of UV (254 nm), or with mitomycin C (10 µg ml⁻¹) left in the medium for the duration of the experiment, or transfected with ~5 µg of pRSVtat-SVneo (K. Valerie, unpublished; part of a cDNA clone of HIV which has the *tat* gene²¹, as a 356 bp *Sal*I-*Bam*HI fragment, under control of the RSV LTR²²). The transfected cells were left untreated or UV irradiated ~20 h after transfection. Similarly, E3-cat HeLa cells (clone 27-1; N. Jones, unpublished) were UV irradiated and processed the same as the HIV-cat HeLa cells. The cells were harvested ~20 h after UV/mitomycin C treatment (~40 h after transfection), and CAT activities determined²⁴. Between 5-15 µg of total protein was used per reaction, including 0.084 µCi of [¹⁴C]-chloramphenicol and 0.63 mM of acetyl-CoA. The reactions were incubated at 37°C, extracted, spotted onto silica plates, developed in chloroform: methanol (95:5), and autoradiographed. To quantitate the CAT activity the radioactive spots were cut out and counted in a liquid scintillation counter. All reactions were linear except the CAT activity obtained after the combined *tat* transfection and UV treatment.

transiently expressed in only a fraction of the cells being analysed. This makes it difficult to assess the extent to which the combined effects of these two different agents are truly being monitored.

We also measured the levels of CAT in the HIV-cat cell line at various times after UV irradiation, and after irradiation with different UV doses. We found that CAT activity could first be detected between 2 and 4 h, reaching maximum levels between 20 and 24 h, and then decreasing from 24 to 48 h after irradiation (Fig. 1b), and that a UV dose >6 J m⁻² was required to induce the HIV LTR (Fig. 1c). Mitomycin C concentrations >1 µg ml⁻¹ were required to induce CAT activity in the HIV-cat cells (data not shown). The UV-induced time course of CAT in HIV-cat cells is very similar to *tat*-mediated induction of CAT activity, which is first detected between 2 and 3 h after polyethylene glycol-facilitated protoplast fusion of HIV-cat HeLa cells with *Escherichia coli* cells producing *tat* protein¹⁵ (data not shown).

The mechanism by which *tat* functions to enhance HIV LTR-directed gene expression is not known and there remains controversy as to whether it exerts its primary effects at the transcriptional and/or post-transcriptional level¹⁰⁻¹⁴. In an attempt to understand how UV treatment enhances LTR-directed gene expression, we first studied the effects of α -amanitin on UV induction in the HIV-cat HeLa cell line. As shown in Fig. 2a, α -amanitin effectively blocked the UV induction of CAT activity, suggesting that RNA polymerase II-directed transcription is required for the effect. Moreover, direct measurement of

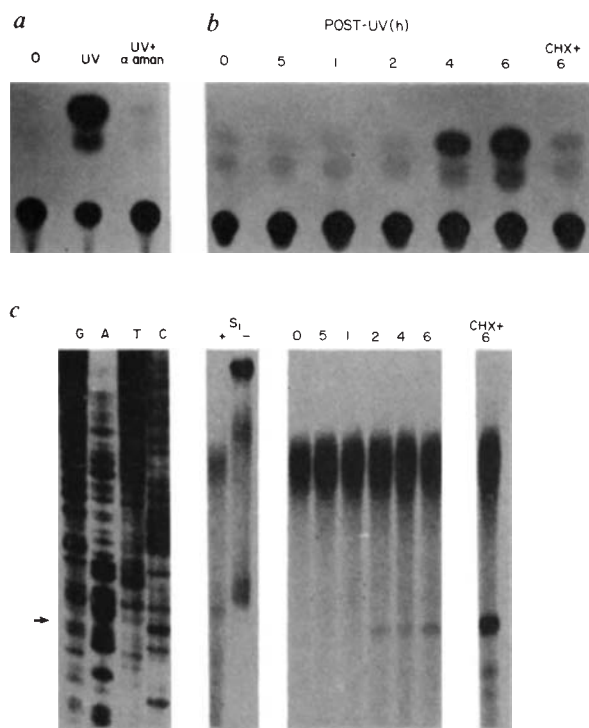


Fig. 2 *a*, Inhibition of UV-induced CAT activity by α -amanitin. HIV-cat HeLa cells were pre-incubated for 1 h with $10 \mu\text{g ml}^{-1}$ of α -amanitin (Sigma) prior to UV irradiation (10 J m^{-2}). We previously determined that α -amanitin concentrations of $>5 \mu\text{g ml}^{-1}$ were needed to eliminate the induction of CAT activity (data not shown). The cells were incubated for 19 h in medium containing α -amanitin, harvested, and CAT activity determined as described in the legend to Fig. 1. *b*, *c*, CAT activities and RNA steady state levels after UV irradiation. HIV-cat HeLa cells were UV irradiated, harvested after the indicated times, and CAT activities/steady state levels of *cat* mRNA determined. CHX, cycloheximide-treated cells; G, A, T, C, sequencing reactions; reactions without any RNA and with (+), or without (-) S₁ nuclease. The arrow indicates the start of transcription.

Methods. Total RNA was isolated from HIV-cat HeLa cells by the guanidinium isothiocyanate/hot phenol method²⁵ at various time points after UV (10 J m^{-2}) treatment. Cells from one 100-mm dish ($\sim 10^7$ cells) were used for each time point. One tenth of the cells were saved for parallel CAT assays. One dish was supplemented with $50 \mu\text{g ml}^{-1}$ cycloheximide (which eliminated $>96\%$ of CAT activity) immediately after UV irradiation. When cycloheximide was added 1 h prior to UV irradiation a similar increase of steady state *cat* mRNA was observed as shown in this figure (data not shown). The amount of *cat* mRNA was semi-quantitated by S₁ nuclease protection analysis²⁵, using a 453 nucleotide long uniformly ^{32}P -labelled single-stranded DNA probe. This complementary probe was obtained by primer extension of a 21-mer (starting at the -1 position from the initiation codon of the *cat* gene) oligonucleotide on a single-stranded M13mp19-HIVcatBGH(+) template. The partially double-stranded M13 molecule was then digested with EcoRV, which cuts 340 bp 5' of the transcriptional start-site in the HIV LTR, and isolated on a urea/polyacrylamide gel. After hybridization of the radioactive probe to the RNA (48°C for ~ 12 h), S₁ nuclease (Pharmacia) was added to 400 units ml^{-1} and the reactions incubated at 24°C for 1 h. The mixtures were phenol extracted/ethanol precipitated and loaded on an 8 M urea/5% polyacrylamide sequencing gel. The predicted size for the protected probe fragment is 113 nucleotides. Sequencing reactions²⁶, using the same M13 template and primer, were used as markers.

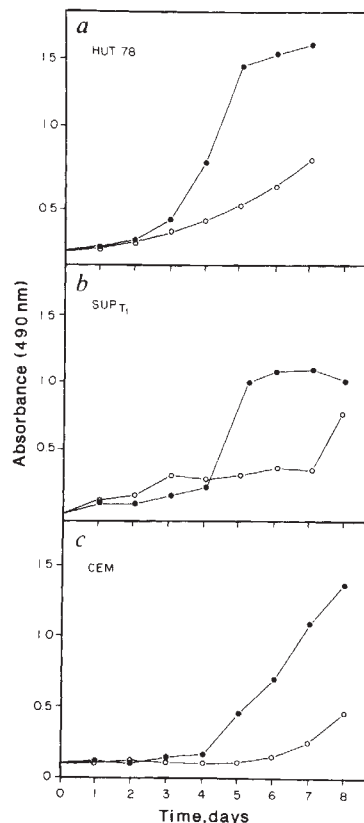


Fig. 3 HIV infection of human T cells after UV treatment. *a*, HUT 78 cells²⁷. *b*, Sup T₁ cells²⁸. *c*, CEM cells²⁹. Untreated cells, (○); UV-treated, (●).

Methods. Approximately 10^7 cells were washed in phosphate-buffered saline and irradiated with 12 J m^{-2} (254 nm) UV or left untreated prior to infection with $100 \mu\text{l}$ of MOLT3/HTLVIIIIB culture supernatant (TCID_{50} per $\text{ml} \sim 3 \times 10^5$ virus; Reed & Muench). Infection was allowed to proceed for 1 h at 37°C , at which point the cells were washed and cultured at 10^6 cells ml^{-1} . Aliquots of cells were collected every day, disrupted with 1% Triton X-100 and assayed for the expression of HIV p24 core antigen by sandwich ELISA, using a monoclonal anti-p24 antibody (DuPont), human HIV-positive serum, and a horseradish peroxidase biotin-streptavidin ELISA kit (Amersham Corp.). The amount of p24 core antigen was quantified by developing with *ortho*-phenylene diamine and measuring the absorbance at 490 nm on an ELISA reader.

cat messenger RNA levels by S₁ nuclease protection analysis indicated that UV treatment results in an increase in steady state levels of *cat* mRNA, which precedes the rise of CAT activity (Fig. 2*b*, *c*). It should be noted that the increased levels of mRNA observed after UV irradiation did not appear to account fully for the dramatic increases in CAT activity. This is similar to what is observed after *trans*-activation of the HIV LTR by *tat*¹⁰⁻¹². Addition of cycloheximide to the cells did not block the UV-induced increase in steady state mRNA levels, although as expected it did block subsequent CAT expression. This indicates that the mRNA increase resulting from UV activation does not require *de novo* protein synthesis, but is mediated by existing cellular factors. We noted that cycloheximide treatment of UV-irradiated cells did result in an accumulation of *cat* mRNA ~ 5 -fold higher than that seen with UV alone (Fig. 2*c*). This additional increase is unlikely to be specific for HIV LTR-directed mRNA, as it has been observed for other transcriptional units^{16,17}.

The fact that UV/mitomycin C treatment had such a dramatic

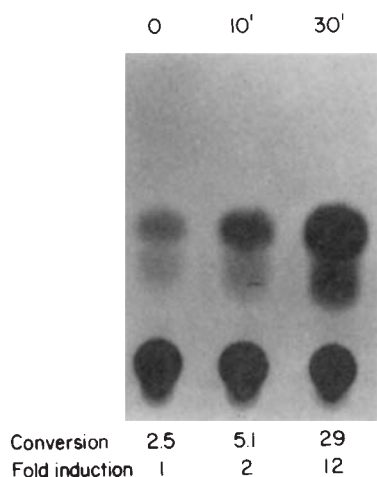


Fig. 4 Activation of the HIV LTR by sun exposure. Semi-confluent HIV-cat HeLa cells in 60 mm tissue culture dishes were washed with phosphate-buffered saline and exposed to sunlight (starting 12:50 pm Philadelphia time, on 18 August, 1987) for the indicated times. The dishes were kept on wet ice, without any cover, facing the sun at an angle of $\sim 90^\circ$. The control dish was kept on ice under a cardboard cover. To avoid drying of the cells 1 ml of buffer was kept in the dish, including the control, which was carefully and repeatedly flicked every minute. The cells were then incubated in medium at 37°C and harvested 23 h after sun exposure. CAT activities, expressed as the % conversion of [^{14}C]-chloramphenicol to acetylated forms, were determined as described in the legend to Fig. 1.

effect on HIV LTR-directed expression in our HIV-cat HeLa cell system suggested that these agents might directly affect HIV development *in vivo*. To test this idea, we UV irradiated three different human T-cell lines, infected them with HIV and assayed for viral growth by monitoring p24 core antigen titres (Fig. 3). Remarkably, virus could be detected significantly earlier (4–5 days) in the irradiated cells, as compared to untreated cells (7–8 days). It is also clear from this data that prior treatment of cells with UV enhances viral growth considerably. This finding is consistent with the effect of UV on HIV LTR gene expression seen in the HIV-cat HeLa cell system and has important practical implications for assaying HIV growth in infected cells since it shows that the time required for detection of virus can be shortened significantly by prior treatment of cells with UV.

Since UV light is a component of sunlight, it was of interest to examine whether this source of radiation could also activate HIV LTR-directed expression. As shown in Fig. 4, exposure of the HIV-cat HeLa cells to sunlight activates the expression of CAT activity. A 30 min sun exposure resulted in a ~ 12 -fold increase in CAT activity, and even after an exposure as short as 10 min some increase in CAT activity (~ 2 -fold) was observed. This result raises the possibility that the virus can be activated by direct sun exposure. The fact that HIV has been isolated from epidermal Langerhans cells of an AIDS patient¹⁸, suggests that this finding may have clinical relevance.

Various chemical and viral agents which have been shown to enhance HIV LTR-directed gene expression result in a wide range of activation spanning from ~ 10 to >100 -fold^{1–6}. The contrast between a low and a high level of activation suggests that these agents work at very different efficiencies and perhaps by different mechanisms. Recent data consistent with this interpretation indicates that the elements responsible for lectin/phorbol ester activation map to the two enhancer regions (-105 to -80) within the LTR^{4–6}. This certainly contrasts to the region responsive to *tat* activation, which maps to an element, TAR, spanning the transcriptional start-site in the LTR and extending

at least 80 bp into the transcribed portion of the leader region¹⁹. Moreover, lectin/phorbol ester activation appears to be synergistic with *tat*, again implying separate mechanisms^{4,5}. In contrast to lectin/phorbol ester induction, UV activates HIV LTR-directed expression at higher levels, similar to those obtained with *tat* itself^{10–15}. The increases we observe in steady state *cat* mRNA levels upon UV induction of HIV-cat HeLa cells are difficult to reconcile with the 150-fold increases seen in gene expression. This is particularly striking as other results indicate that the TAR sequence negatively affects the expression of mRNA *in cis* in the absence of *tat* function (refs 11, 12; H.R., C.D. and M.R., unpublished observations). This suggests that UV irradiation, like *tat*^{10–12}, might exert at least some of its effect at the post-transcriptional level. However, because the element(s) responsive to UV activation has not been identified, it is as yet unclear whether this agent affects the same regulatory pathway(s) as *tat*.

It has been proposed that the *tat* gene product and TAR sequence play an important role in reactivating viral expression from latency^{11–13}. What, then, initiates the process? It is possible that cellular factors induced or activated by a variety of agents are primarily responsible for triggering a cascade of events which include initial expression of *tat* protein followed by highly elevated HIV production. Thus, HIV provirus activation may well be induced by certain stress responses in an analogous way to λ prophage induction in *E. coli* cells^{7,8}. If this is true, an SOS-like stress response could be a general inducer of the HIV LTR. Perhaps the way the *tat* protein exercises its potency, once it has been produced in sufficient amounts, is by maintaining the stress-induced state to assure a highly increased capacity to direct HIV LTR expression. This would further imply that *tat* functions via a cellular mechanism normally involved in stress-related responses in mammalian cells.

We thank S. Arya, D. Pfarr, J. Trill and M. Reff for supplying some of the parent plasmids used in this study, N. Jones for E3-cat HeLa cells, Y.-S. Ho for providing laboratory space and R. Sweet for critical reading of the manuscript. This work was supported in part by an NIH grant (K. V., H. R., C. D. and M. R.), and a grant from Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture (A. D., C. B. and C.T.).

Received 7 December; accepted 14 March 1988.

- Gendelman, H. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9759–9763 (1986).
- Mosca, J. D. *et al. Nature* **325**, 67–70 (1987).
- Bednarek, D. P., Mosca, J. D. & Raj, N. B. *K. J. Virol.* **61**, 1253–1257 (1987).
- Nabel, G. & Baltimore, D. *Nature* **326**, 711–713 (1987).
- Tong-Starksen, S. E., Luciw, P. A., & Peterlin, B. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6845–6849 (1987).
- Kaufman, J. D. *et al. Molec. cell. Biol.* **7**, 3759–3766 (1987).
- Radman, M. *Photochem. Photobiol.* **32**, 823–830 (1980).
- Herrlich, P., Mallick, U., Ponta, H. & Rahmsdorf, H. J. *Human Genet.* **67**, 360–368 (1984).
- Herrlich, P. *et al. Adv. Enzyme Regul.* **25**, 485–504 (1986).
- Rosen, C. A. *et al. Nature* **319**, 555–559 (1986).
- Feinberg, M. B. *et al. Cell* **46**, 807–817 (1986).
- Cullen, B. R. *Cell* **46**, 973–982 (1986).
- Peterlin, B. M., Luciw, P. A., Barr, P. J. & Walker, M. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9734–9738 (1986).
- Wright, C. M., Felher, B. K., Paskalis, H. & Pavulakis, C. *Science* **234**, 988–992 (1986).
- Fischer, A. G. *et al. Nature* **320**, 367–371 (1986).
- Imbra, R. J. & Karin, M. *Nature* **323**, 555–558 (1986).
- Imbra, R. J. & Karin, M. *Molec. cell. Biol.* **7**, 1358–1363 (1987).
- Tschachler, E. *et al. J. invest. Derm.* **88**, 233–237 (1987).
- Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813–823 (1985).
- Southern, P. J. & Berg, P. *J. molec. appl. Genet.* **1**, 327–341 (1982).
- Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69–73 (1985).
- Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6777–6781 (1982).
- Pfarr, D. P., Sathie, G. & Reff, M. E. *DNA* **4**, 461–467 (1985).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
- Maniatis, T., Fritsch, W. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Sanger, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Gazdar, A. F. *et al. Blood* **55**, 409–417 (1980).
- Hecht, F., Morgan, R., Kaiser-McCaw Hecht, B. & Smith, S. D. *Science* **226**, 1445–1447 (1984).
- Foley, G. E. *et al. Cancer* **18**, 522–529 (1965).

An inducible promoter fused to the *period* gene in *Drosophila* conditionally rescues adult *per*-mutant arrhythmicity

John Ewer, Michael Rosbash & Jeffrey C. Hall

Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

The *period* (*per*) gene of *Drosophila melanogaster* is involved in the expression of circadian rhythms of locomotor activity in adult flies¹. Molecular studies of *per* (reviewed in ref. 2) have shown that the transcribed and translated products of this gene are present primarily at the embryonic, pupal and adult stages³⁻⁸. Here we describe experiments with arrhythmic *per* mutants bearing an inducible form of this gene which indicate that strongly rhythmic adult behaviour can be obtained only if *per* expression is induced in the adult, independent of its history of expression earlier in development. Thus *per*-mutant locomotor-activity phenotypes seem not to result from abnormalities in the development of neural structures or in physiological processes that may be required at pre-adult stages for the expression of this circadian rhythm. Moreover, the action of *per* after light:dark cycle entrainment seems to be sufficient for activity rhythms to be exhibited in constant darkness; this suggests further that the *per* product is

required only during the time that the rhythmic behaviour is being manifested. Our strategy used a heat-shock gene promoter fused to *per* coding sequences to obtain conditional gene expression. Heat-shock promoter-driven genes have previously been used to study the mode of action and tissue specificity of a variety of *Drosophila* genes⁹⁻¹⁶; our experiments on circadian rhythms demonstrate the use of such gene constructions for the temporal manipulation of genes whose phenotypes, behavioural and otherwise, affect whole organisms.

To construct a conditional *per*-gene *Drosophila*, arrhythmic *per*⁰¹ hosts were transformed with a heat-shock protein-70 promoter-*per* fusion gene (*hsp-per*, see legend to Fig. 1) using P element-mediated transformation¹⁷. The adult locomotor activity of these transformants was assayed at high or low temperature following their exposure to different heat-treatments during development and adulthood.

Three independent transformant lines were obtained in which rhythmicity could be induced only by exposing the animals to high temperature. Thus, when transformants from these three lines were raised at a low temperature (18 °C), transferred to 29 °C and monitored at this high temperature under constant (free-running) conditions, their locomotor activities were clearly organized in bouts. These bouts occurred regularly throughout the run (Fig. 1a), in a manner similar to that seen in flies expressing rhythmic *per* alleles. A statistical analysis (χ^2 periodogram) of these data (as in ref. 18) indicated that a significant period could be detected in these records (Fig. 1a), although it was longer than that of wild-type rhythms. In contrast, when

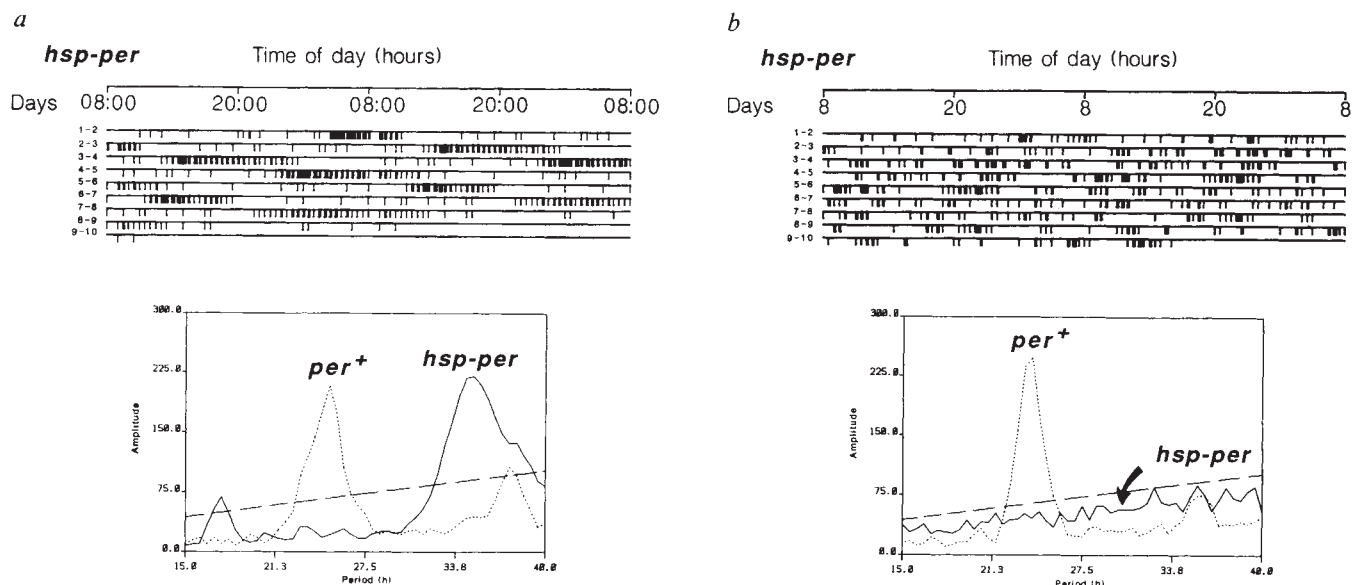


Fig. 1 Conditional locomotor activity rhythm of *hsp-per* transformants. *a*, Top, locomotor activity over time (actogram) of an *hsp-per* transformant (from line S9, Fig. 2) after a heat treatment regime (see Fig. 2a ii). Rhythmic locomotor activity is clearly apparent. (In this actogram the data have been 'double-plotted', such that days 1-2 of freerun are displayed on the top horizontal line, days 2-3 on the next line, and so on.) Bottom, statistical analysis (χ^2 periodogram¹⁸) of activity of *hsp-per* flies (—) from the actogram shown above, and of *per*⁺ flies (---, actogram not shown). Statistically significant periods are defined as those values that exceed the 95% confidence limit indicated in the periodograms (—). Note that the (significant) period is longer than the one derived from *per*⁺ flies under these conditions. *b*, Top, actogram of *hsp-per* transformant (from line S1) not subjected to heat-treatment. The activity pattern appears to be the same as that of straight *per*⁰¹ animals^{2,18,23,26}, both by inspection and by periodogram analysis (below). Bottom, periodogram corresponding to the actogram shown above (—) compared to periodogram of a *per*⁺ fly raised and tested at 18 °C (---, actogram not shown).

Methods. The *hsp70-per* fusion gene was constructed by ligating a 450 base pair (bp) sequence containing the *hsp70* promoter^{9,10} to a 6.8 kilobase (kb) *Xba*I-*Eco*RI *per* locus DNA fragment that includes 300 bp of the first intron and complete coding region^{28,29}. The *hsp70* fragment contains 200 bp of untranslated sequences downstream from the cap site^{9,10,13}. The *hsp-per* fusion was inserted into a Carnegie 20 P-element vector³⁰ which carries the *rosy* (*ry*⁺) gene as marker, and used to transform *per*⁰¹; *ry*⁻ embryos (as described in ref. 23). Of the six independent insert-bearing lines so generated, two were seldom rhythmic at any temperature and one was rhythmic at high and low temperatures; for these reasons, only the lines noted in Fig. 2 were analysed extensively. In all behavioural experiments, the *hsp-per* flies tested were the *ry*⁺ male segregants from a cross of a single *ry*⁺ male from a given transformant line with *per*⁰¹; *ry*⁻ females; thus, these progeny carry one dose of a given *hsp-per* insert. In this experiment, the animals were reared at 18 °C; at eclosion they were either transferred to 29 °C (*a*) or kept at the low temperature (*b*). One or two days later they were entrained for 5 days in a 12 h:12 h light:dark cycle and then transferred to constant darkness (free-run) for monitoring of locomotor activity at 29 °C or at 18 °C. As controls, 12 *per*⁰¹ mutants were heat-treated as adults, and all except one (period, 39.5 h) were arrhythmic. In contrast, all 16 *per*⁺ controls thus tested were rhythmic (period, 24.7 ± 0.1 h). Behavioural data were collected and analysed as described¹⁸.

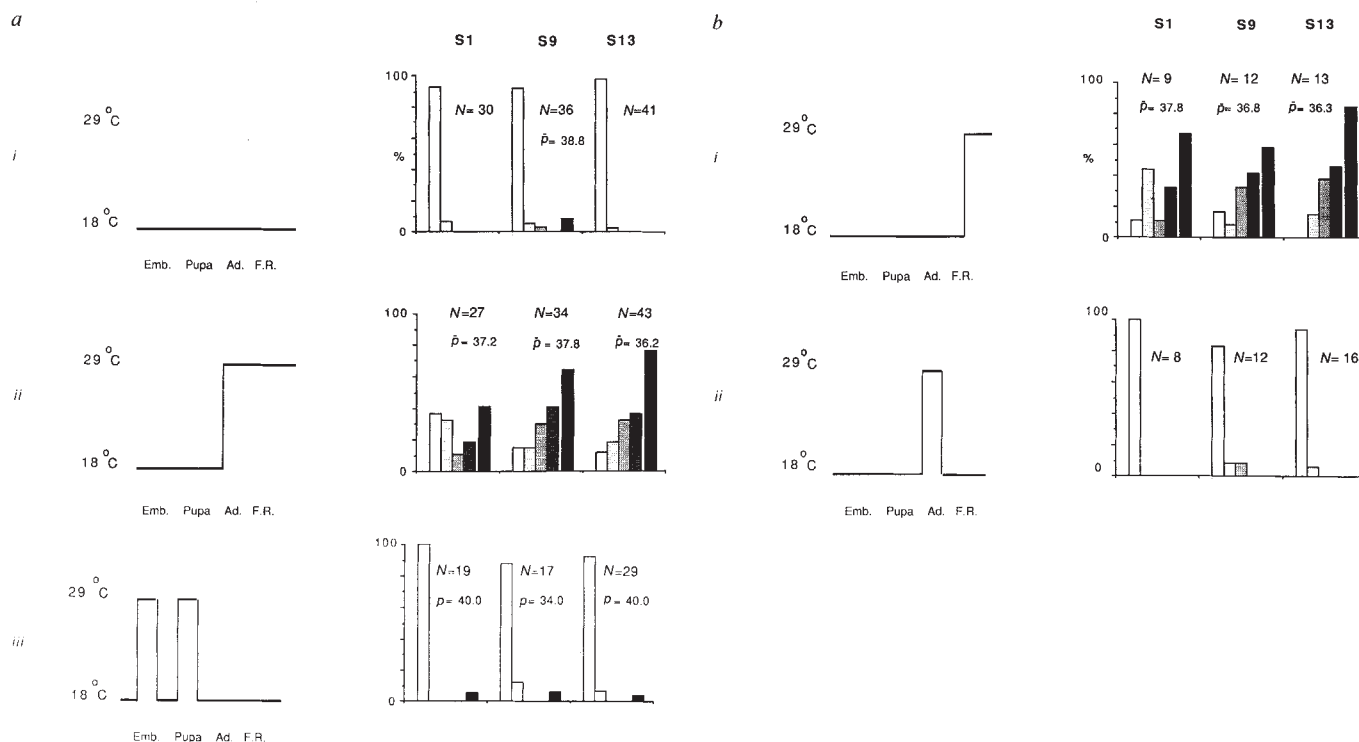


Fig. 2 Summaries of the *per*^{ol} rescue in *hsp-per* transformants under different temperature regimes. For each of three independent transformant lines (S1, S9 and S13, which have a single transduced insert on chromosome 3, 2 and 3 respectively) the extent of rescue has been quantified by indicating the percentage of the total number of flies tested (*N*) whose locomotor activity was indistinguishable from that of *per*^{ol} animals (□), or which exhibited bouts of activity that were barely discernible (▨), clearly apparent but without fixed periodicity (▩) or both apparent and periodic (■) (as in Fig. 1*a*, top). These four categories sum to 100%. The percentage of animals whose periodograms had a statistically significant period¹⁸, and the average period in hours ($\bar{p}$ or p , if only one animal had a statistically significant period) is also indicated (■). The temperatures to which the transformants were exposed during embryonic (Emb.), pupal (Pupa) and adult (Ad.) stages, as well as during the freerun (F.R.) period are indicated on the left. *a*, Extent of rescue obtained when transfer from 18 °C to 29 °C was effected at the adult (ii) or embryonic and pupal stages (iii). The phenotypes of the transformants were different from those maintained continuously at 18 °C (i) only if they were heat treated as adults (ii). The extent of rescue was not increased if transformants were exposed to 29 °C at pre-adult stages in addition to the adult one. Exposure to standard heat-shock conditions (30 min at 37 °C, as in ref. 31) did not effect any visible rescue, nor did it increase the extent of rescue obtained when delivered in addition to transfers from 18 °C to 29 °C. *b*, The extent of rescue obtained when flies were transferred from 18 °C to 29 °C during free-run (i) was similar to that obtained when this transfer was done before the test period (ii). No rescue could be elicited by an exposure of adults to 29 °C that ended before the test period (ii).

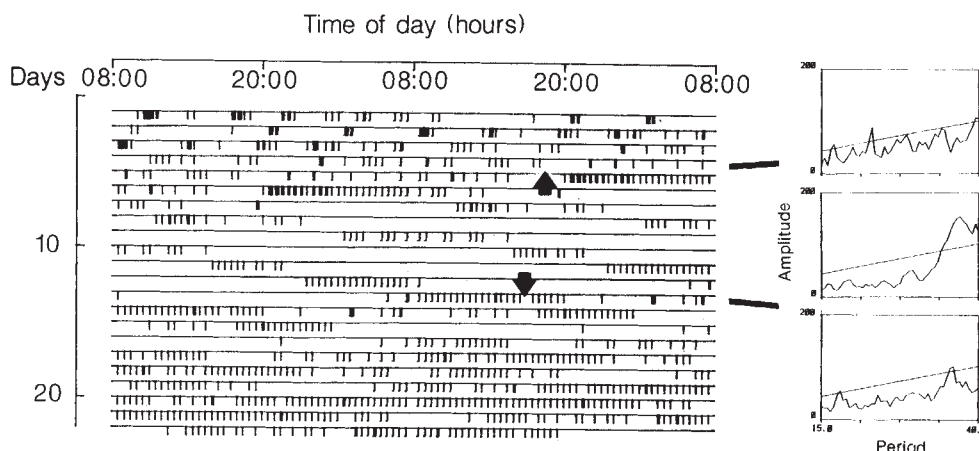
raised and tested at the low temperature, the locomotor activities of transformants from the same lines were characteristic of those of the background (*per*^{ol}) arrhythmic genotype (Fig. 1*b*). The results were essentially the same in experiments using all three of the transformed lines (Fig. 2) and thus indicate that these *hsp-per* transformants behave as conditional *per* mutants.

To determine the stages at which *per* expression is required for the manifestation of a rhythmic adult phenotype we monitored, under free-running conditions, the locomotor activity of *hsp-per* transformants that had been transferred from 18 °C to 29 °C at different stages of the life cycle. These stages corresponded to those during which *per* is expressed in wild-type flies³⁻⁸. The results, summarized in Fig. 2*a*, indicate that *per* expression in the adult is both necessary and sufficient for a rhythmic phenotype.

Genes driven by the *hsp70* promoter in transformants can be weakly active under non-heat-shock conditions^{12,16,19}. Thus, a low level of *per* messenger RNA transcribed from the hybrid gene could be present when the animals are kept at 18 °C. Although this level is clearly not sufficient for the expression of rhythmicity in adults (Fig. 2*a*), it could be enough to effect some rescue of functions that may be necessary, in addition to adult ones, for the rescue of arrhythmicity. If this were so, one could expect that increasing expression of the *hsp-per* fusion at pre-adult stages would lead to stronger rhythmicity (that is higher penetrance and with periodicities closer to wild-type, see below). However, neither exposures of pupae + adults, nor of embryos + pupae + adults, to 29 °C improved the extent of rescue.

The locomotor activity rhythm of transformants during high temperature treatment was not as robust as in the wild-type^{1,12,18}. First, the periods of rescued transformants were consistently longer than wild-type (Fig. 1*a* and Fig. 2*aii*). We do not believe that this is caused by a lack of *per* gene activity at pre-adult stages (see above), but instead interpret the >24 h periods as resulting from low *hsp-per* expression in the relevant adult cells (which are unknown, though see ref. 7). Previous genetic and molecular evidence suggests an inverse relationship between *per* activity and the period of circadian rhythms²⁰⁻²². Hybridization *in situ* to sections of transformants heat-shocked for 30 min at 37 °C revealed, as expected^{9,11}, widespread hybridization in all tissues (data not shown). Thus, although the overall levels of *hsp70* promoter-driven *per* RNA induced by heat treatment were higher than those of *per* RNA found in the wild-type (data not shown), the widespread expression of this fusion gene could mean that the actual levels per cell were insufficient for wild-type function. The poor rhythmicity is probably not due to some debilitating effect caused by the ectopic *per* expression in these transformants because *hsp-per* bearing chromosomes, when put in a *per*⁺ background, had no effect on the wild-type phenotype (each of six such flies was rhythmic, $\bar{p} = 23.9 \pm 0.3$ h). A second problem with the phenotype of the transformants is that only 40–80% of the heat-treated adults were strongly rhythmic (Fig. 2*aii*). We do not have an explanation for this incomplete penetrance, but the same kind of phenomenon has been reported for other *per* transformants^{23,18}, as well as for *Drosophila* transformed with other *hsp70* promoter-driven genes^{13,15}.

Fig. 3 A rhythmic phenotype can be both elicited and interrupted during free-run. Left, actogram of the locomotor activity of a transformant (line S9) raised and entrained at 18 °C, then tested at this temperature during the first 6 days of freerun. On the sixth day (↑) the temperature was raised to 29 °C, and maintained at this temperature for the next 8 days. Note that a rhythmic pattern (significant by a χ^2 periodogram, $p = 37.0$ hr), was induced rapidly in a previously arrhythmic record (see Fig. 2bii for a summary of results). Rhythmicity was essentially lost when the temperature was returned to 18 °C on day 14 of free-run (↓). The total percentages of records that were statistically arrhythmic after the return from 29 °C to 18 °C were: 92 (line S1, $N = 13$), 90 (S9, $N = 19$) and 94 (S13, $N = 18$); these include results obtained in two separate experiments. Right, χ^2 periodograms from the actogram data collected before the transfer from 18 °C to 29 °C (top), during free-run at 29 °C (centre), and following the return to 18 °C. In eight *per^{ol}* controls tested at the same time as the transformants, no rhythmicity was induced by the transfer to 29 °C. In eight *per⁺* animals tested, this transfer, which occurred after several days of free-run (see above) and at a phase corresponding to 6 h after noon (referring to the entrainment period), induced immediate phase delays of ~6 h in the locomotor activity rhythms. Nevertheless, the records for such wild-type adults were significantly rhythmic before, during and after their 8-day heat treatments. Because of the effect that this temperature shift had on the activity of the wild-type pacemaker, we do not know at present whether the rhythmicity evoked in the transformants was due solely to the induction of *per* expression.



We now provide evidence that the *per* product is required in the adult only during the manifestation of rhythmicity: indeed, as exemplified in Fig. 3, the transfer of our *hsp-per* flies to the high temperature during free-run rapidly induced the appearance of a rhythmic phenotype. The extent of rescue obtained (Fig. 2b) was similar to that observed when transformants were exposed to high temperature starting several days before their locomotor activity began to be monitored (Fig. 2a). Therefore, not only is the role of *per* in regulating the circadian activity pacemaker apparently not developmental; it also seems not to be required for the interpretation and storage of light:dark entrainment cues.

Circadian rhythmicity does not persist in the absence of *hsp-per* induction. No significant rescue was effected if transformants were exposed to 29 °C during entrainment only (Fig. 2bii). Likewise, if the heat-induction was discontinued during the free-run period, previously rhythmic animals rapidly became arrhythmic (Fig. 3). These data imply that the effective half-life of the *per* product from the *hsp70*-driven *per* gene is short (as in ref. 12), and are consistent with it playing, at the adult stage, a purely physiological role in the expression of rhythmic adult activity.

We conclude that mutations at the *per* locus do not disrupt the development of the pacemaker that underlies circadian rhythms of locomotor activity in *Drosophila* adults. There are other phenotypes of *per* mutants, in larvae^{6,24,25} and in pupae^{1,26}, which may result from *per*-related developmental and morphological defects. In adults, however *per* mutations do not lead to gross anatomical defects (although see ref. 27); in fact the total absence of this gene leaves the flies thoroughly viable, albeit arrhythmic^{18,26}.

We suggest that in the adult the *per* gene product plays only a physiological role in the manifestation of locomotor rhythmicity. Indeed, inducing the expression of this gene after development has been completed is both necessary and sufficient

for this phenotype to be induced. Furthermore, the rapidity with which rhythms can be turned on or off during free-running conditions seems to argue that *per* regulates the ongoing operations of oscillator functions, as opposed to being involved in the construction of the fly's circadian pacemaker.

We thank Qiang Yu and Melanie Hamblen for technical assistance and Gary Struhl for providing an *hsp70* promoter-containing vector. We appreciate comments on the manuscript from Kathy Siwicki, Hildur Colot and John Nambu. This work was supported by a US NIH grant to M.R. and J.C.H.

Received 19 January; accepted 14 March 1988.

1. Konopka, R. J. & Benzer, S. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2112-2116 (1971).
2. Hall, J. C. & Rosbash, M. *J. Biol. Rhythms* **2**, 153-178 (1987).
3. Reddy, P. *et al. Cell* **38**, 701-710 (1984).
4. Bargiello, T. A. & Young, M. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2142-2146 (1984).
5. James, A. J., Ewer, J., Reddy, P., Hall, J. C. & Rosbash, M. *EMBO J.* **5**, 2313-2320 (1986).
6. Bargiello, T. A. *et al. Nature* **328**, 686-691 (1987).
7. Siwicki, K. K., Eastman, C., Petersen, G., Rosbash, M. & Hall, J. C. *Neuron* (in the press).
8. Liu, X., Lorenz, L., Yu, Q., Hall, J. C. & Rosbash, M. *Genes Dev.* **2**, 228-238 (1988).
9. Lis, J. T., Simon, J. A. & Sutton, C. A. *Cell* **35**, 403-410 (1983).
10. Dudler, R. & Travers, A. A. *Cell* **38**, 391-398 (1984).
11. Bonner, J. J., Parks, C., Parker-Thornburg, J., Mortin, M. A. & Pelham, H. R. B. *Cell* **37**, 979-991 (1984).
12. Steller, H. & Pirrotta, V. *EMBO J.* **4**, 3765-3772 (1985).
13. Struhl, G. *Nature* **318**, 677-680 (1985).
14. Steller, H. & Pirrotta, V. *Molec. cell. Biol.* **6**, 1640-1649 (1986).
15. Schneuwly, S., Klemenz, R. & Gehring, W. J. *Nature* **325**, 816-818 (1987).
16. Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. *Cell* **50**, 739-747 (1987).
17. Rubin, G. M. *Trends Neurosci.* **8**, 231-233 (1985).
18. Hamblen, M. *et al. J. Neurogenet.* **3**, 249-291 (1986).
19. Klemenz, R., Hultmark, D. & Gehring, W. J. *EMBO J.* **4**, 2053-2060 (1985).
20. Smith, R. F. & Konopka, R. J. *Molec. gen. Genet.* **185**, 30-36 (1982).
21. Coté, G. & Brody, S. J. *J. theoret. Biol.* **121**, 487-503 (1986).
22. Baylies, M. K., Bargiello, T. A., Jackson, F. R. & Young, M. W. *Nature* **326**, 390-392 (1987).
23. Zehring, W. A. *et al. Cell* **39**, 369-376 (1984).
24. Livingstone, M. S. *Neurosci. Abstr.* **7**, 351 (1981).
25. Weitzel, G. & Rensing, L. *J. Comp. Physiol.* **143**, 229-235 (1981).
26. Smith, R. F. & Konopka, R. J. *Molec. gen. Genet.* **183**, 243-251 (1981).
27. Konopka, R. J. & Wells, S. J. *J. Neurobiol.* **11**, 411-415 (1980).
28. Jackson, R. F., Bargiello, T. A., Yun, S.-H. & Young, M. W. *Nature* **320**, 185-188 (1986).
29. Citri, Y. *et al. Nature* **326**, 42-47 (1987).
30. Rubin, G. M. & Spradling, A. C. *Nucleic Acids Res.* **11**, 6341-6351 (1983).
31. Ashburner, M. & Bonner, J. J. *Cell* **17**, 241-254 (1979).

Identification of an altered splice site in Ashkenazi Tay-Sachs disease

E. Arpaia*, A. Dumbrille-Ross*, T. Maler*, K. Neote*, M. Tropak*, C. Troxel†, J. L. Stirling‡, J. S. Pitts‡, B. Bapat*, A. M. Lamhonwah*, D. J. Mahuran*, S. M. Schuster†, J. T. R. Clarke§, J. A. Lowden‡ & R. A. Gravel*¶

* Research Institute, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

† Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA

‡ MRC Human Genetics Disease Research Group, King's College London, Kensington Campus, London W8 7AH, UK

§ Department of Pediatrics, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

‖ HSC Research Development Corporation, 88 Elm Street, Toronto, Ontario M5G 1X8, Canada

Tay-Sachs disease is an autosomal recessive genetic disorder resulting from mutation of the *HEXA* gene encoding the α -subunit of the lysosomal enzyme, β -N-acetylhexosaminidase A (ref. 1). A relatively high frequency of carriers (1/27) of a lethal, infantile form of the disease is found in the Ashkenazi Jewish population, but it is not yet evident whether this has resulted from a founder effect and random genetic drift or from a selective advantage of heterozygotes². We have identified a single-base mutation in a cloned fragment of the *HEXA* gene from an Ashkenazi Jewish patient. This change, the substitution of a C for G in the first nucleotide of intron 12 is expected to result in defective splicing of the messenger RNA³. A test for the mutant allele based on amplification of DNA by the 'polymerase chain reaction'⁴ and cleavage of a *DdeI* restriction site generated by the mutation revealed that this case and two other cases of the Ashkenazi, infantile form of Tay-Sachs disease are heterozygous for two different mutations. The occurrence of multiple mutant alleles warrants further examination of the selective advantage hypothesis.

The mutant *HEXA* gene was cloned from a skin fibroblast line, GM2968, derived from a patient with the Ashkenazi, infantile form of Tay-Sachs disease. GM2968 was previously shown to have normal *HEXA* DNA by Southern blot analysis and undetectable mRNA for the α subunit^{5,6}. Our strategy for the identification of the mutation was to sequence the exons and splice junctions, and both ends of the gene, as these were the most likely sites for mutation that would affect mRNA synthesis or stability⁷. The sequence analysis of a cloned fragment of the gene revealed a G to C substitution in the first nucleotide of the 5' boundary of intron 12 (Fig. 1). Significantly, the first two nucleotides, GT, in introns have been shown to be invariant and are obligatory for the proper splicing of mRNA³. Mutations at this position in splice junctions disrupt normal splicing and result in the production of aberrant mRNA species which can have severely reduced stability⁸. Examples are found in β -thalassaemia⁸, phenylketonuria⁹ and haemophilia B¹⁰.

A fortuitous result of the mutation is the generation of a *DdeI* restriction site which is absent in normal DNA (Fig. 1). We took advantage of this site to detect the mutant allele in samples of DNA extracted from the cells of Tay-Sachs patients or fetuses and their parents. This was done by amplifying a 151 base-pair (bp) sequence of DNA containing the mutation using the polymerase chain reaction (Fig. 2), cleaving with *DdeI* and examining the products by polyacrylamide gel electrophoresis (Fig. 3). This strategy was initially evaluated using a plasmid containing part of the normal *HEXA* gene (Fig. 3, lanes 1-4). After amplification and *DdeI* cleavage, it yielded a 120-bp fragment

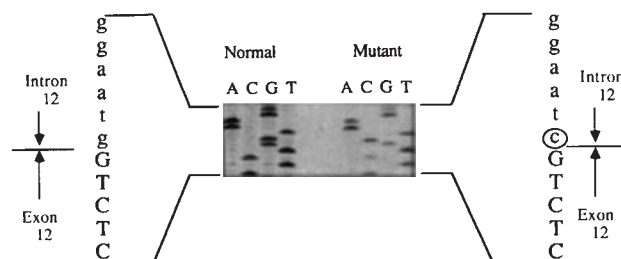


Fig. 1 Nucleotide sequence of normal and mutant DNA at donor splice junction of exon 12. The normal *HEXA* gene was cloned from a library of a partial *HaeIII*-*AluI* digest of human fetal liver DNA¹⁷. The mutant *HEXA* gene was cloned from fibroblast line GM2968 (NIGMS Human Genetic Mutant Cell Repository, Camden, New Jersey). The library was made as a partial *Sau3A* digest in Lambda Dash (Stratagene). Sequencing was done on a 5.8 kilobase (kb) *EcoRI* fragment cloned into Bluescript (Stratagene). Both strands were sequenced by the Sanger dideoxy-chain-termination method¹⁸ using double-stranded DNA as template. Oligonucleotide primers were made on a Vega Coder 300 DNA synthesizer at the University of Nebraska. The mutation (circled) is shown as the first nucleotide of intron 12.

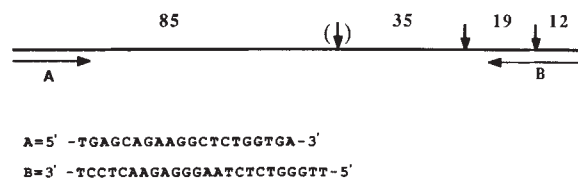


Fig. 2 Structure of the amplified sequence surrounding the mutation site (arrow in parentheses). Positions of *DdeI* sites are indicated by the vertical arrows. The oligonucleotide primers used for DNA amplification are shown as horizontal arrows labelled A and B with the nucleotide sequences as shown. The numbers above the solid line are the fragment sizes (bp) generated by *DdeI* cleavage. The total length is 151 bp.

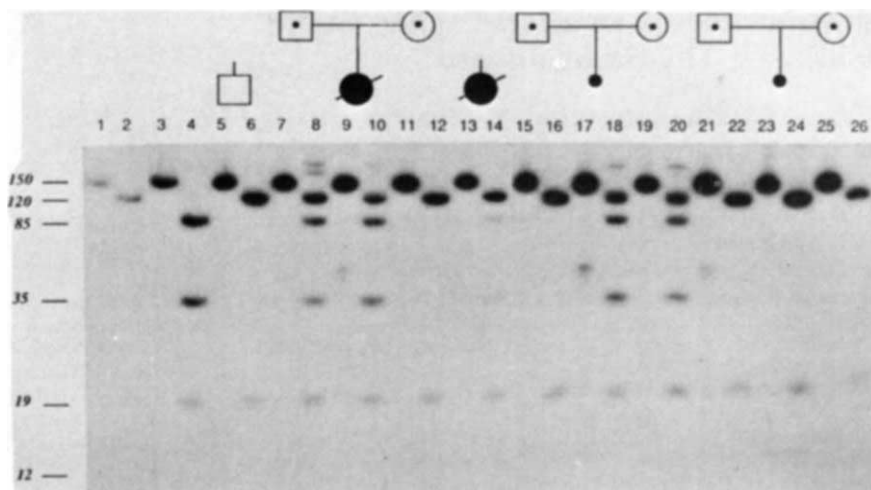
and the expected smaller fragments (Fig. 3, lane 2). In contrast, the 120-bp band was absent in DNA amplified from a plasmid containing the mutant DNA but was replaced by two fragments of 85 and 35 bp (lane 4). These observations show that the relevant sequence was amplified and confirm the presence of the novel *DdeI* site in the cloned DNA.

Amplification and cleavage of DNA from normal cells also produced the 120-bp fragment (Fig. 3, lane 6). When genomic DNA from GM2968 was amplified and cleaved with *DdeI*, the 85- and 35-bp fragments were detected (lane 10), confirming the presence of the mutation site. But the 120-bp fragment was also seen. The appearance of this fragment could not be due to incomplete digestion by *DdeI*, as its occurrence required cleavage at a control site. Therefore, the result can only be explained if the patient is heterozygous for the splice mutation. Cleavage of amplified DNA from cells of the father (lane 8) gave the same pattern as the patient, whereas the mother showed only the 120-bp fragment (lane 12), indicating that the mutation came from the father. These results show that the patient must be heterozygous for two different mutations, the splice mutation identified by the *DdeI* site and another mutation as yet undefined at the molecular level.

We next examined DNA from affected fetuses and parents of two additional Ashkenazi Jewish families. In one case, DNA amplified from the mutant cells and cut with *DdeI* gave the diagnostic bands of 120, 85 and 35 bp as before (Fig. 3, lane

¶ To whom correspondence should be addressed.

Fig. 3 Polyacrylamide gel electrophoresis of amplified DNA surrounding the mutation site at exon 12 in cloned plasmid and cellular DNA. Samples are arranged in pairs: odd numbers show amplified, uncleaved DNA; even numbers to the right are the same samples cleaved with *DdeI*. Symbols at the top show individual pedigrees: open symbol, normal; dotted symbol, carrier; filled large symbol, affected; filled small symbol, affected terminated pregnancy. Numbers to the left are sizes of DNA fragments determined from nucleotide sequence and confirmed with size markers (not shown). Lanes 1-4 contain amplified plasmid DNA, and lanes 5-26 are amplified genomic DNA. Lanes 1-2, cloned normal DNA; lanes 3-4, cloned GM2968; lanes 5-6, normal DNA; lanes 7-8, GM3051; lanes 9-10, GM2968; lanes 11-12, GM3052, lanes 13-14, GM515A; lanes 15-26 are from families seen at the Hospital for Sick Children.



Method. Plasmid (0.2 µg) or genomic DNA (5-10 µg) were amplified for 27 cycles (about one millionfold) by the polymerase chain reaction using Taq polymerase (Perkin-Elmer), 1.25 U per reaction, as in ref. 19. Conditions for each cycle were 2 min for denaturation of DNA at 95 °C, 3 min for annealing the DNA and primers at 50 °C, and 4 min for primer extension at 72 °C. The final extension was for 11 min. α -[³²P]dCTP (New England Nuclear), 10 µCi, was included in each reaction to radiolabel the newly synthesized DNA. An aliquot (about 10% of total) was digested with 9 U of *DdeI* (New England Biolabs) for 3 hours at 37 °C in a total volume of 20 µl. The complete reaction mix and an equivalent amount of uncleaved DNA were applied to a 20% polyacrylamide/0.5% agarose gel and electrophoresis was carried out at 40 mA in Tris-borate-EDTA buffer²⁰. The gels were exposed without drying to X-ray film overnight.

18), with the mutation segregating through the mother (lanes 16 and 20 versus 18). The other family did not show the described mutation pattern at all, in either the affected fetus or the parents (lanes 22, 24 and 26). Also, the cell line of one more affected individual, GM515A, also had the mutation in single dose (lane 14). Thus, of eight *HEXA* alleles examined from four patients with the Ashkenazi, infantile form of Tay-Sachs disease, three have the *DdeI* restriction site diagnostic of the G to C substitution.

These results do not rule out the possibility that we have identified an unusual polymorphism rather than a mutation causing Tay-Sachs disease. We therefore tested for the allele in DNA samples from known carriers¹¹ and normal Jewish subjects obtained from the Tay-Sachs Screening Service at the Hospital for Sick Children. The splice mutation was detected in two of 14 confirmed carriers and none of 43 normal Jewish individuals. Thus, this mutation accounts for over 20% of the Tay-Sachs alleles that we have examined in Ashkenazi Jewish patients and carriers. We propose, therefore, that the nucleotide change identified in GM2968 is responsible for *HEXA* gene dysfunction in the Ashkenazi, infantile form of Tay-Sachs disease. Although we cannot formally exclude a polymorphism in strong linkage disequilibrium with an unidentified Tay-Sachs mutation as an alternative explanation, the incompatibility of the mutant sequence with normal splicing is a compelling argument in favour of mutation.

The high carrier frequency of Tay-Sachs disease among Ashkenazim has generally been ascribed to a founder effect¹² with the expectation that a single mutant allele would be identified to account for it. There has been considerable debate whether genetic drift and migration or a selective advantage of carriers can best explain the spread of the Tay-Sachs allele^{13,14}. Our finding that at least two alleles for infantile Tay-Sachs disease must be segregating in this population changes the emphasis of debate about the origins of the high carrier frequency. It will now be fascinating to see whether there is a single mutant allele in the majority of carriers or a greater

number of mutant alleles. The infantile form of Tay-Sachs disease also occurs in high frequency among French Canadians in parts of Quebec¹⁵. Although clinically indistinguishable from the Ashkenazi infantile disease, it is caused by a mutation distinct from the one described here. Myerowitz and Hogikyan¹⁶ have shown that it is due to a deletion of a portion of the 5' end of the *HEXA* gene. We found that the deletion mutation does not account for the unidentified alleles in the four patients examined in this study.

We thank M. A. Skomorowski for providing Tay-Sachs screening samples, Nancy Rooke-Allen for assistance with cell culture, E. Lower and Perkin-Elmer Canada for assistance with Taq polymerase, and B. and R. Glickman for helpful discussion. This work was supported by the Medical Research Council of Canada, the National Institutes of Health and the Wellcome Trust.

Received 2 February; accepted 21 March 1988.

- O'Brien, J. S. in *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S.) 945-969 (McGraw-Hill, New York, 1983).
- Kaback, M. M. in *Lysosomes and Lysosomal Storage Diseases* (eds Callahan, J. W. & Lowden, J. A.) 331-342 (Raven, New York, 1981).
- Rogers, J. H. *Int. Rev. Cytol.* **93**, 187-231 (1985).
- Saiki, R. K. *et al. Science* **230**, 1350-1354 (1985).
- Myerowitz, R. & Hogikyan, N. D. *Science* **232**, 1646-1648 (1986).
- Korneluk, R. G. *et al. J. Biol. Chem.* **261**, 8407-8413 (1986).
- Proia, R. L. & Sorovio, E. *J. Biol. Chem.* **262**, 5677-5681 (1987).
- Treisman, R., Orkin, S. H. & Maniatis, T. *Nature* **302**, 591-596 (1983).
- DiLella, A. G., Marvit, J., Lidsky, A. S., Güttler, F. & Woo, S. L. *Nature* **322**, 799-803 (1986).
- Rees, D. J., Rizza, C. R. & Brownlee, G. G. *Nature* **316**, 643-645 (1985).
- Kaback, M. M., Shapiro, L. J., Hirsch, P. & Roy, C. *Prog. Clin. Biol. Res.* **18**, 267-279 (1977).
- Chase, G. A. & McKusick, V. A. *Amer. J. Hum. Genet.* **24**, 339-340 (1972).
- Wagener, D., Cavalli-Sforza, L. L. & Barakat, R. *Amer. J. Hum. Genet.* **30**, 262-270 (1978).
- Myrianthopoulos, N. C. & Melnick, M. *Prog. Clin. Biol. Res.* **18**, 95-106 (1977).
- Andermann, E., Scriver, C. R., Wolfe, L. S., Dansky, L. & Andermann, F. *Prog. Clin. Biol. Res.* **18**, 161-188 (1977).
- Myerowitz, R. & Hogikyan, D. *J. Biol. Chem.* **262**, 15396-15399 (1987).
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157-1174 (1978).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Wong, C. *et al. Nature* **330**, 384-386 (1987).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma

Buzzy Morse, Paul G. Rotherg*, Victoria J. South, John M. Spandorfer* & Susan M. Astrin

Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania 19111, USA

The proto-oncogene *c-myc* is the cellular homologue of the transforming sequence carried by the avian myelocytomatosis virus MC29. A growing body of evidence implicates structural and functional alterations in and around proto-oncogenes such as *c-myc* in tumorigenesis. Here we report that comparison of the structure of *myc* from a ductal adenocarcinoma of the breast and from normal breast tissue of the same patient (Sc) revealed a tumour-specific rearrangement of one *myc* locus and amplification of the other *myc* locus. (For *myc* reviews see refs 1-4; for *myc* involvement in breast neoplasia see refs 5-7.) Within the second intron of the rearranged locus was a non-*myc* sequence with nearly complete homology to a long interspersed repetitive element (a LINE-1 sequence or L1). In this case, the L1 sequence has functioned as a mobile genetic element to produce a somatic mutation.

Figure 1a shows a series of Southern blots of DNA from normal and from carcinoma tissue of patient Sc. Figure 1b shows a restriction map and indicates the hybridizing DNA fragments from the normal *myc* locus. With tumour DNA (middle and right panels of Fig. 1a), the *myc* bands of 19, 13, 2.9, 2.3, 1.9 and 1.5 kilobases (kb), expected for an unrearranged *myc* locus, were unusually intense, suggesting a tumour-specific amplification of one *myc* allele. Averaging over several Southern blot experiments in which tumour and normal DNAs were compared on the same blots, and normalizing for another single-copy gene (*int-1*) (data not shown), we determined a level of *myc* gene amplification of approximately 10-fold for tumour DNA. Therefore, the tumour-specific, unrearranged, amplified *myc* locus was present at ~20 copies per cell.

Figure 1a also demonstrates a tumour-specific rearrangement as indicated by DNA bands seen with *Sac*I (the 4.2 kb fragment), *Eco*RI (the 3.8 kb fragment), and *Pvu*II/*Eco*RI (the 3.8 kb fragment) using an exon 3 probe (middle panel), and with *Eco*RI (the 16 kb fragment) using an exon 2 probe (right panel). Other Southern blots (data not shown) demonstrate that the 10 kb *Pvu*II band (exon 3 probe, middle panel) and the 5.5 kb *Kpn*I band (exon 2 probe, right panel) shown in Fig. 1a are also tumour specific. Averaging over several Southern blots (data not shown), we estimate the tumour-specific rearranged band to be present at one copy per cell. The ability to detect the tumour-specific rearranged 3.8 kb *Eco*RI fragment with exon 3 probe (middle panel), but not with exon 2 probe (right panel) suggests that sequences containing a new *Eco*RI site have been introduced between the *myc* coding exons 2 and 3 (generating the 3.8 kb *Eco*RI fragment containing the *myc* third exon, Fig. 1a, middle panel).

We cloned the smaller of the two rearranged *Eco*RI fragments (see Fig. 1a). Three primary clones were detected with a *myc* exon 3 probe. Each contained a 3.8 kb *Eco*RI insert fragment with an identical restriction pattern, which also co-migrated with the 3.8 kb *Eco*RI fragment seen on Southern blots using Sc's tumour DNA (data not shown). One insert fragment was subcloned into the plasmid pBR322. The resulting subclone was designated 3.8pScR1 (see Fig. 2), and represents only a part of

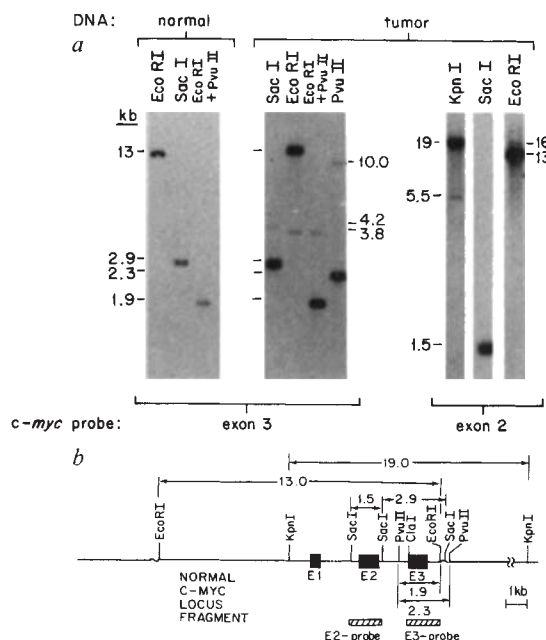


Fig. 1 Southern blot analysis of *myc* in Sc, a breast carcinoma patient. *a*, Southern blots of normal and tumour DNAs digested with the restriction enzymes shown and hybridized with the probes indicated at the bottom of the panel. Band sizes are in kb. *b*, Restriction map of the normal *myc* locus to demonstrate the fragments expected in the Southern blot hybridizations shown in *a*. The normal-sized but amplified bands in the tumour lanes of *a* (middle and right panels; 19 kb *Kpn*I, 13 kb *Eco*RI, 2.9 kb *Sac*I, 2.3 kb *Pvu*II, 1.9 kb *Eco*RI/*Pvu*II, and 1.5 kb *Sac*I fragments) are shown on the normal *myc* fragment. Sizes are in kb; E1 = exon 1, E2 = exon 2, E3 = exon 3. E2 and E3 probe fragments are indicated. **Methods.** Normal tissue and a 4 × 6 cm poorly-differentiated infiltrating ductal-adenocarcinoma were removed from a 47-year-old woman (Sc) by surgeons at the Fox Chase Cancer Center. Tissues were immediately flash-frozen in liquid nitrogen and stored at -70 °C until used. Standard techniques were used for all DNA preparations from mammalian cells¹². Aliquots (10 µg) of high molecular weight, normal (nonmalignant breast) and tumour DNAs were restricted with the enzymes shown, according to manufacturer's (New England Biolabs) instructions. Electrophoresis¹³, DNA transfer, hybridization and washing¹⁴ were as described. Two human *myc* probes were derived from the 13 kb *Eco*RI clone (AhR1-15) of the human *myc* gene as described previously¹³. The second exon probe (E2-probe) was a 1.5 kb *Sac*I fragment, and the third exon probe (E3-probe) was a 1.4 kb *Cl*aI-*Eco*RI fragment (see Fig. 1b).

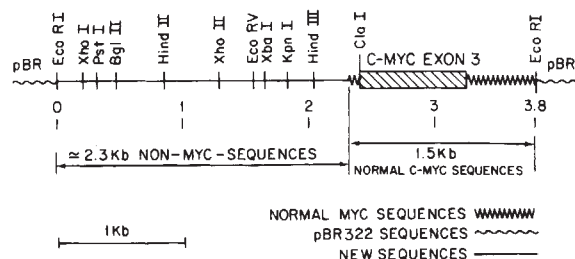


Fig. 2 Partial restriction map of the 3.8pScR1 subclone in plasmid pBR322. The subclone contains ~2.3 kb of non-*myc* sequences (shown) and ~1.5 kb of normal *myc* sequences. **Methods.** Standard techniques were used for DNA preparations from *Escherichia coli* (HB101 and LE392), lambda cloning (EMBL4, Stratagene, Inc.), and restriction enzyme analysis and mapping¹². The restriction endonuclease sites shown are those from which the DNA was sequenced (see Fig. 3).

* Present addresses: Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263, USA (P.G.R.) and Jefferson Medical College of Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA (J.M.S.).

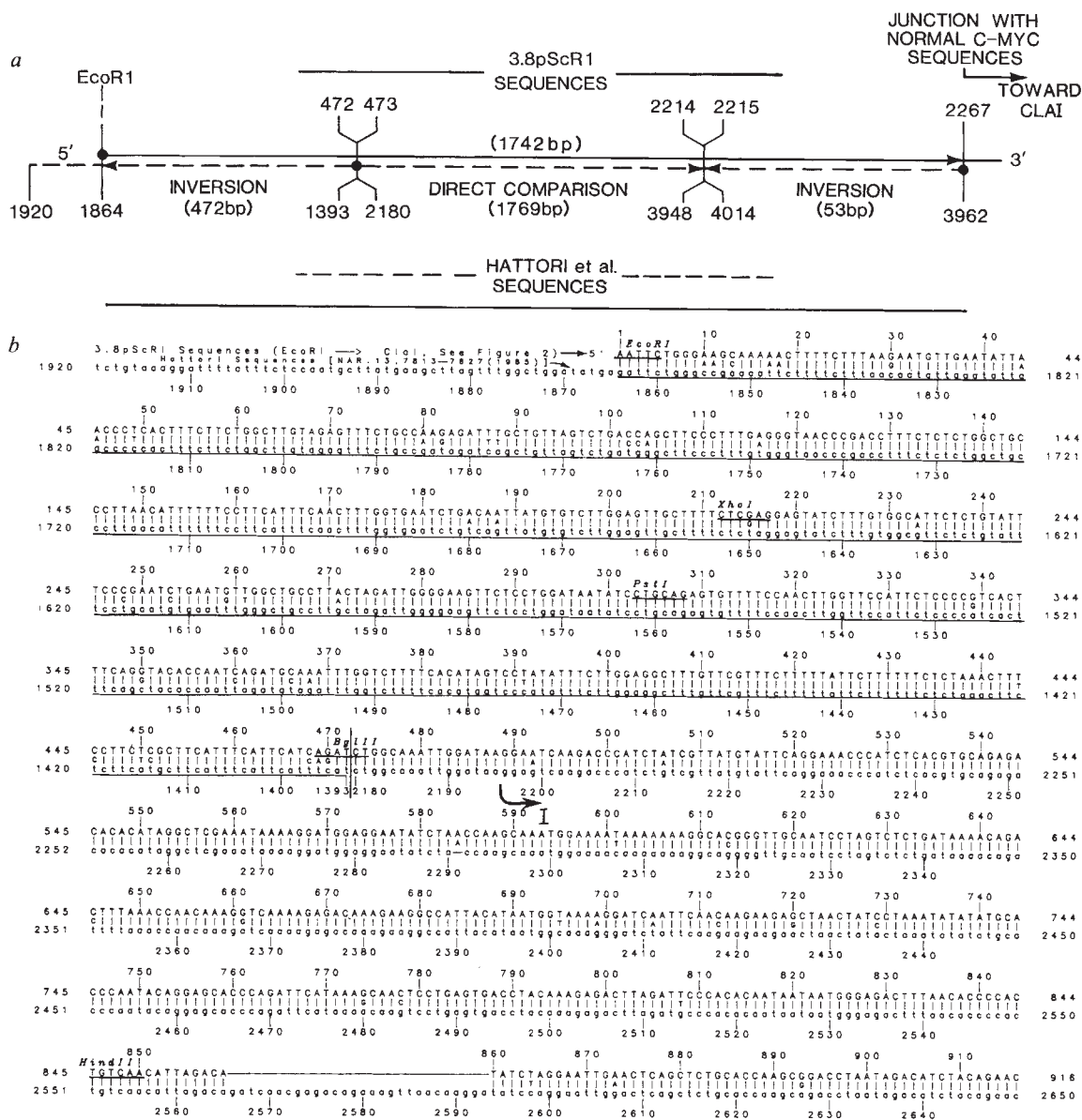
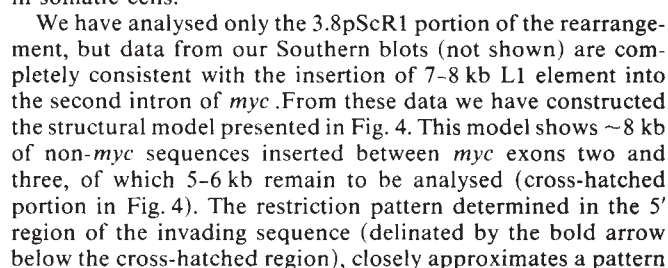


Fig. 3 Comparison of the non-*myc* 3.8pScR1 subclone sequences and portions of the L1 sequence⁸. *a*, Schematic diagram (not to scale) of the sequence comparison in *b*. The 3.8pScR1 sequences are represented as a solid-lined arrow in one direction only, the Hattori *et al.* L1 sequence as broken-lined arrows. Arrows pointing in the same direction indicate a direct comparison; arrows in opposite directions indicate an inverted comparison (we took the complement of the Hattori *et al.* sequences for comparison). The discrepancy in nucleotide numbers seen in the middle portion of the diagram reflect the 28 bp deletion in the 3.8pScR1 sequences (nucleotides 2,566–2,593). *b*, Nucleotide sequence comparison. The 3.8pScR1 sequences (nucleotides 1–2,267) start at the 5' *EcoRI* site and end near the *ClaI* site as shown in Fig. 2. The Hattori *et al.* L1 sequences are numbered as they were presented in the original paper⁸. The 3.8pScR1 sequences are on top, the L1 sequences below. Restriction endonuclease sites used during sequencing are noted above the underlined hexanucleotide recognition sequences. Horizontal dashes represent deletions. The vertical dashes between sequences indicate nucleotide homology. In cases of non-homology, the base determined for that position within the 3.8pScR1 sequence is shown between the sequences. Blocks of the L1 sequence underlined in ▶

the rearranged locus. The remainder is contained in the 16kb *EcoRI* fragment shown in Fig. 1a right panel and in the model diagram of Fig. 4. Repeated efforts failed to identify unique sequences within the non-*myc* portion of the 3.8pScR1 subclone, as all tested fragments yield only smears when used to probe human genomic Southern blots (data not shown). The results indicate that a highly repetitive element had become contiguous with *myc*'s third exon.

The restriction sites indicated in Fig. 2 were used for sequencing the 3.8pScR1 subclone. The new sequences (~2,300 base pairs (bp) shown in Fig. 2 from the 5' *EcoRI* site to near the 3' *ClaI* site) are presented in Fig. 3, along with a comparison to a previously identified full-length L1 element⁸ (for L1 element

reviews, see 9–11). The data show that the 3'-most breakpoint within the *myc* second intron occurred 87 bp 5' to the *myc* third exon, keeping intact the second intron/third exon acceptor splice site. Figure 3a diagrams the regions of homology. In Fig. 3b, the 3.8pScR1 sequence is presented in full (nucleotides 1–2,267), but only those portions of the L1 sequence with homology to the 3.8pScR1 sequence are shown. Nucleotides 1–472 in our sequence are very similar (39 bp differences) to an inverse complement of the L1 sequence (nts 1,864–1,393, underlined in Fig. 3b). The next 1,742 bp are homologous with the L1 sequence (nts 473–2,214, our sequence; nts 2,180–3,948, L1 sequence), except for a 28 bp deletion relative to the L1 sequence (nts 2,566–2,593) and 123 isolated single bp differences. Finally,



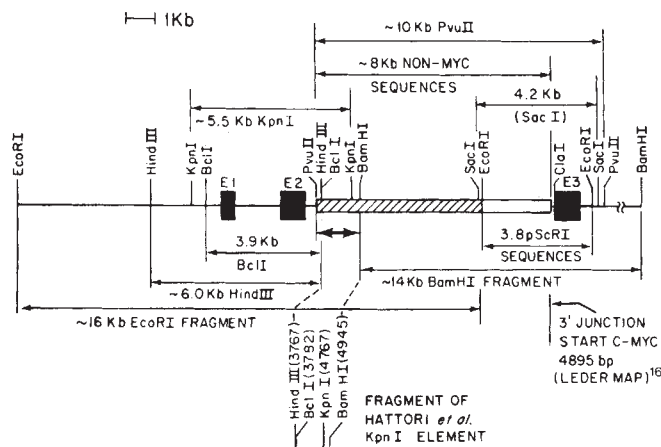


Fig. 4 Structural model of rearranged *myc* locus containing an inserted L1 element. The diagram shows a blocked area of ~8 kb of non-*myc* sequences inserted between *myc* coding exons 2 and 3, and is based upon preliminary Southern data as explained in the text. The open portion of the blocked area, labelled as 3.8pScR1, is described in the text. The cross-hatched portion of the blocked area represents the remainder of the non-*myc* sequences we have not yet cloned or analysed. The presumed origin of the lighter intensity, tumour-specific, rearranged *myc* bands (the 16 kb *EcoRI*, 10 kb *PvuII*, 4.2 kb *SacI*, 3.8 kb *EcoRI*, and 3.8 kb *EcoRI/PvuII* bands) seen in Fig. 1a are shown in the model. Other restriction fragments sizes (the 14 kb *BamHI*, 6 kb *HindIII*, 5.5 kb *KpnI*, and 3.9 kb *BclI* fragments) have been seen on many Southern blots (data not shown) and have been used to map the 5' end of the proposed insert. The partial restriction map of a fragment taken from the L1 element (numbered as in the original paper⁸) is provided for comparison with the restriction endonuclease pattern we determined in the 5' region of the insert.

identified within the full-length L1 element⁸. Therefore, we postulate that the entire disrupting sequence is an L1 element which has inserted between *myc* coding exons 2 and 3. The relatively large size of this L1 may be a reflection of either the presence of novel L1 sequences, the existence of non-L1 sequences in the body of the element, or of internal reiterations. Alternatively, due to inherent errors in determining DNA molecular size using agarose gels, this element may in fact be smaller (in the normal range 6–7 kb). Cloning of the entire element will resolve these possibilities, and allow us to study the complete structure of this interesting L1 element.

We thank Drs Maxine Singer, Steven Potter, Jerry Freed and Alfred Knudson for criticisms of this manuscript, Dr Potter for a *KpnA* clone, Dr Michael Erisman for *myc* probes, Dr Vinayaka Prasad, Dr Munir Khan and Vidya Jonnalagadda for discussions, Mark Shlomchik, Robert Kleinfeld and Anita Cywinski for sequencing assistance, Rich Bomze, Willie Barr, Harold Ott, Karen Trush, Joe Hurley and Clara Brett for graphics work, Bob Stodola, Dr Frank Manion and Paul Marcolina for help in computer sequence analysis and Secretarial Services for preparation of this manuscript.

This work was supported by grants an Appropriation from the Commonwealth of Pennsylvania, P.G.R. was a Special Fellow of the Leukemia Society of America. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00304.

Received 1 February; accepted 17 March 1988.

1. Croce, C. M. *Cancer Res.* **46**, 6019–6023 (1986).
2. Bishop, J. M. *Science* **235**, 305–311 (1987).
3. Bishop, J. M. *A. Rev. Biochem.* **52**, 301–354 (1983).
4. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1707–1711 (1983).
5. Escot, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4834–4838 (1986).
6. Kozbor, D. & Croce, C. M. *Cancer Res.* **44**, 438–441 (1984).
7. Leder, A., Pattengale, P. K., Kuo, A., Stewart, T. A. & Leder, P. *Cell* **45**, 485–495 (1986).
8. Hattori, M., Hidaka, S. & Sakaki, Y. *Nucleic Acids Res.* **13**, 7813–7827 (1985).

9. Fanning, T. G. & Singer, M. F. *Biochim. biophys. Acta* **910**, 203–212 (1987).
10. Singer, M. F. & Skowronski, J. *Trends biochem. Sci.* **10**, 119–122 (1985).
11. Skowronski, J. & Singer, M. F. *Cold Spring Harb. Symp. quant. Biol.* **51**, 457–464 (1986).
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (eds) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
13. Rothberg, P. G., Erisman, M. D., Diehl, R. E., Rovigatti, U. G. & Astrin, S. M. *Molec. cell Biol.* **4**, 1096–1103 (1984).
14. Erisman, M. D. *et al. Molec. cell Biol.* **5**, 1969–1976 (1985).
15. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
16. Battey, J. *et al. Cell* **34**, 779–787 (1983).
17. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).
18. Isono, K. *Nucleic Acids Res.* **10**, 85–89 (1982).

Heavy-chain binding protein recognizes aberrant polypeptides translocated *in vitro*

C. K. Kassenbrock, P. D. Garcia, P. Walter & R. B. Kelly

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448, USA

Immunoglobulin heavy-chain binding protein (BiP, GRP-78) associates tightly in the endoplasmic reticulum (ER) with newly synthesized proteins that are incompletely assembled, have mutant structures, or are incorrectly glycosylated^{1–3}. The function of BiP has been suggested to be to prevent secretion of incorrectly folded or incompletely assembled proteins^{2,3}, to promote folding or assembly of proteins³, or to solubilize protein aggregates within the ER lumen⁴. Here we examine the interaction of BiP with newly synthesized polypeptides in an *in vitro* protein translation–translocation system. We find that BiP forms tight complexes with nonglycosylated yeast invertase and incorrectly disulphide-bonded prolactin, but does not associate detectably with either glycosylated invertase or correctly disulphide-bonded prolactin. Moreover, BiP associates detectably only with completed chains of prolactin, not with chains undergoing synthesis. We conclude that BiP recognizes and binds with high affinity *in vitro* to aberrantly folded or aberrantly glycosylated polypeptides, but not to all nascent chains as they are folding.

Immunoglobulin heavy-chain binding protein (BiP) has been shown to be identical to the glucose regulated protein of relative molecular mass (M_r) 78,000 (GRP-78) and is a member of the evolutionarily conserved HSP-70 family of heat-shock related proteins⁴. Manipulations which inhibit N-linked glycosylation of proteins induce BiP synthesis^{5,6}. Furthermore, BiP shows increased binding to immunoglobulin heavy chains², influenza haemagglutinin³, and tissue plasminogen activator⁷ when glycosylation is inhibited with tunicamycin. Exposure of cells to anoxia⁸, amino acid analogues⁹ and sulphhydryl-reducing agents¹⁰, which do not specifically affect glycoprotein synthesis, also induce BiP synthesis. In addition, BiP binds to mutant but correctly glycosylated proteins³. A plausible explanation of these observations is that BiP recognizes incorrectly folded proteins.

To examine whether BiP associates only with aberrant proteins or also plays a role in the folding of newly synthesized proteins in the ER, we used a coupled protein translation–translocation system consisting of a wheat germ translation extract supplemented with dog pancreas rough microsomes. Addition of messenger RNA that encodes a secreted protein leads to the efficient targeting and cotranslational translocation of the protein into the lumen of the microsomes¹¹. As BiP is an abundant component of dog pancreas microsomes (C.K.K. and R.B.K., unpublished data), an interaction of BiP with newly synthesized proteins can be detected by precipitating BiP complexes from the lysed microsomes using a BiP-specific antibody.

Yeast invertase, a highly glycosylated protein, was used to examine the effects of altered glycosylation on association with BiP. To inhibit the addition of N-linked oligosaccharide to the

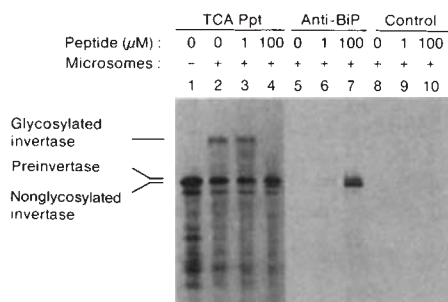


Fig. 1 BiP associates with incorrectly glycosylated invertase but not with native invertase. Composite fluorograph of an SDS polyacrylamide gel. Lanes 1-4, the trichloroacetic acid (TCA)-precipitable polypeptides synthesized *in vitro* from preinvertase mRNA under different reaction conditions: lane 1, products synthesized in the absence of dog pancreas microsomes; lanes 2-4, products synthesized in the presence of microsomes and the indicated concentrations of the tripeptide Ac-Asn-Tyr-Thr-NH₂; lanes 5-10, immune precipitations of aliquots of the same reactions, with a monoclonal antibody against BiP (lanes 5-7) or an unrelated control monoclonal antibody (lanes 8-10). Lanes 5-10 were exposed to film ~8 times longer than lanes 1-4.

Methods. Preinvertase mRNA was transcribed and capped *in vitro* as described²¹ from the plasmid pS5 (provided by J. Ngsee and M. Smith). At time zero mRNA was added to wheat germ extract-dog pancreas microsome translation-translocation reactions (refs 22, 23) containing the tripeptide Ac-Asn-Tyr-Thr-NH₂ at the indicated concentrations. After 1 h at 26 °C the reactions were terminated by adding apyrase for 2 min at 37 °C (100 units ml⁻¹ final concentration) to deplete ATP. Samples were then placed on ice and microsomes were lysed with Nikkol (0.1% final concentration). Aliquots of the reaction were withdrawn and precipitated with TCA (15% final concentration) or monoclonal antibodies. Samples for immune precipitation were incubated with monoclonal culture supernatants directed against either BiP² or an unrelated control antigen at 4 °C overnight. Protein A Sepharose was then added for 45 min at room temperature and the resultant immune precipitates were washed with 0.1% Nikkol in HEPES buffered saline. All precipitates were solubilized with SDS and analysed by electrophoresis on 10-18% gradient SDS polyacrylamide gels. Samples were visualized by fluorography with sodium salicylate. The tripeptide was synthesized at the Biomolecular Resource Center at USC. HPLC analysis showed a purity of over 99%. Protease protection experiments (not shown) indicated that the tripeptide did not interfere with the translocation process.

protein we used the tripeptide Ac-Asn-Tyr-Thr-NH₂ which competes with nascent polypeptide chains as an acceptor for carbohydrate addition¹² (Fig. 1, compare lane 2 with lane 4). Signal peptide cleavage occurs normally, allowing translocated (nonglycosylated) chains to be distinguished from peptides which have not been translocated into the microsomes. Immune precipitation of BiP using a monoclonal antibody² reveals an association of BiP with a proportion (23%) of the incorrectly glycosylated invertase chains (Fig. 1, lane 7). No interaction of BiP with correctly glycosylated invertase can be detected under these conditions (Fig. 1, lanes 5 and 6). We conclude that BiP can specifically recognize and bind tightly to aberrantly glycosylated peptides *in vitro*. In Fig. 1 the incorrectly glycosylated invertase appears as a doublet (lane 7). Although the upper band comigrates with preinvertase, it cannot be the untranslocated precursor as association with BiP only occurs when glycosylation is blocked. Instead, the two bands probably represent unglycosylated and partially glycosylated invertase.

We next asked whether changes in protein structure other than altered glycosylation could lead to complex formation with BiP. To examine whether BiP can recognize changes in protein folding in the absence of altered amino acid or carbohydrate composition we used bovine prolactin as a model protein. Prolactin is a single polypeptide chain of 199 amino acids that

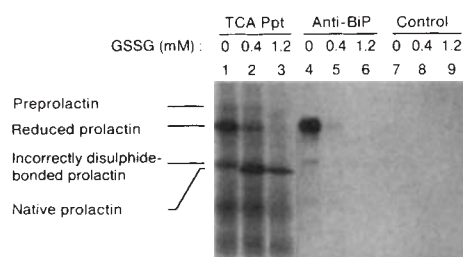


Fig. 2 BiP associates with reduced or incorrectly disulphide-bonded prolactin but not with native prolactin. Composite fluorograph of a non-reducing SDS polyacrylamide gel: lanes 1-3, TCA precipitable polypeptides synthesized *in vitro* from preprolactin mRNA in the presence of microsomes and the indicated concentrations of oxidized glutathione; lanes 4-9, immune precipitations of aliquots of the same reactions, with a monoclonal antibody against BiP (lanes 4-6) or an unrelated control monoclonal antibody (lanes 7-9). Lanes 4-9 were exposed to film ~7.5 times longer than lanes 1-3.

Methods. Preprolactin mRNA was transcribed and capped *in vitro* as described²¹. Translation-translocation reactions (refs 22, 23) containing oxidized glutathione (GSSG) in place of the tripeptide at the indicated concentrations are described in Fig. 2. After termination, samples were placed on ice and microsomes were lysed with Nikkol (0.1% final concentration) and alkylated with iodoacetamide (20 mM final concentration). Aliquots of the reaction were withdrawn and precipitated with TCA or monoclonal antibodies as described in Fig. 1. All precipitates were solubilized with SDS and analysed by electrophoresis on 10-18% gradient SDS polyacrylamide gels under non-reducing conditions, similar to the procedures of Kaderbhai and Austen¹⁴. Under these conditions, native prolactin has the mobility indicated. It runs with reduced prolactin on a reducing gel (not shown). 'Reduced prolactin' migrates identically on both reducing and non-reducing gels. The band marked 'Incorrectly disulphide-bonded prolactin' migrates as reduced prolactin on a reducing gel but is slower than native prolactin on a non-reducing gel. Samples were visualized by fluorography with sodium salicylate.

contains three intra-chain disulphide bonds and is not glycosylated. The *in vitro* protein translation system normally contains dithiothreitol in sufficient concentration to inhibit disulphide-bond formation¹³ but they can be made to form *in vitro* by altering the redox potential using oxidized glutathione^{13,14}. The extent of disulphide bond formation can be monitored on non-reducing SDS-PAGE, where disulphide-bonded prolactin migrates faster than reduced prolactin, presumably due to its more compact configuration¹⁴. Figure 2 shows the results of an experiment in which prolactin has been synthesized with varying degrees of disulphide bonding. Lane 1 shows the products synthesized in the absence of oxidized glutathione. Most of the prolactin migrates with the mobility of reduced prolactin although 22% migrates more rapidly (but not with the mobility of native prolactin) showing that it contains some disulphide bonds. In lanes 2 and 3, where oxidized glutathione is added to the reaction, a new species with the mobility of native prolactin appears. Immune precipitation of BiP reveals the specific association of BiP with a fraction (7%) of the reduced prolactin chains (lane 4). In addition a smaller amount (about 1%) of the prolactin that is incorrectly disulphide-bonded co-precipitates with BiP. No correctly disulphide-bonded prolactin co-precipitates with BiP. This experiment clearly shows that BiP can discriminate between nonglycosylated polypeptides of identical amino acid composition and can form a tight association with polypeptides which are aberrant only in folding and disulphide bonding.

There is evidence that proteins fold, form disulphide bonds, and even oligomerize, before synthesis is complete^{15,16}. If BiP were to facilitate one of these steps, an interaction of BiP with nascent polypeptide chains might be expected. To detect association of BiP with nascent chains, the rate of synthesis was slowed

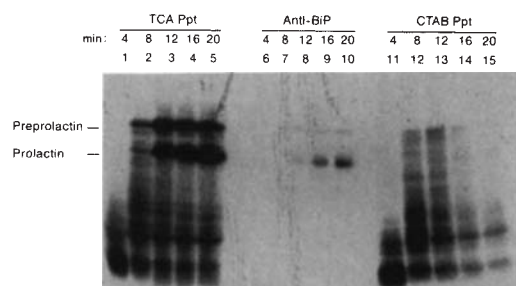


Fig. 3 BiP associates with mature prolactin but not with nascent chains. Fluorograph of a reducing SDS polyacrylamide gel: lanes 1–5, TCA precipitations of polypeptides synthesized at different times *in vitro* in a synchronized translation using preprolactin mRNA in the presence of microsomes; lanes 6–10, immune precipitations of aliquots from the same time points with a monoclonal antibody against BiP; lanes 11–15, aliquots that have been precipitated with CTAB which precipitates nascent chains covalently bound to tRNA¹⁷.

Methods. At time zero preprolactin mRNA was added to a wheat germ extract/dog pancreas microsome translation-translocation system at 22 °C. After 4 min 7-methyl guanosine (5 mM final concentration) and edeine (5 μ M final concentration) were added to block further initiation of peptide synthesis. At the indicated times aliquots of the reaction were withdrawn and treated with glycerol kinase (250 μ g ml⁻¹ final concentration) and glycerol (5% final concentration) for 1 min at 22 °C to deplete ATP and samples were then placed on ice. Each time point was split into three fractions for precipitation with TCA, CTAB²⁴, or with monoclonal antibody to BiP (as in Fig. 1). Precipitates with CTAB and TCA were treated with 1 M Tris base and boiled in SDS to remove the tRNA. All precipitates were solubilized with SDS, reduced with 2-mercaptoethanol, and analysed by SDS-PAGE followed by fluorography as described in Fig. 1.

by performing the reaction at 22 °C instead of 26 °C, and the *in vitro* translation system was synchronized by adding preprolactin mRNA at time zero. Four minutes later, the initiation of new peptide chains was inhibited with edeine and 7-methyl guanosine. Those chains already initiated continued to elongate. At various times during the course of the reaction aliquots were withdrawn and analysed in three ways: acid precipitation of the lysate revealed the total products synthesized; precipitation with the detergent hexadecyltrimethyl ammonium bromide (CTAB) allowed the visualization of nascent chains still covalently bound to transfer RNA¹⁷; and immune precipitation of BiP revealed products that were tightly bound to BiP (Fig. 3). As the redox conditions of this reaction prevent correct disulphide-bond formation, BiP is found tightly associated with a fraction of the full-length prolactin chains (lanes 8, 9 and 10) consistent with the data from non-reducing gels (Fig. 2). None of the nascent chains observed in lanes 11–15 co-precipitated with BiP (lanes 6–10) however. A tight association of BiP with nascent prolactin chains is not detectable even after considerably longer exposure of the gel (not shown).

Not all the preprolactin synthesized *in vitro* is translocated across the membrane and processed by signal peptidase. Densitometry of the gels (Fig. 3) shows that the ratio of prolactin to preprolactin is 3 to 1. Preprolactin is also found in the immunoprecipitates generated with anti-BiP antibodies. This is unlikely, however, to represent specific association of BiP with the extraluminal preprolactin after microsome lysis as in control experiments (not shown) preprolactin (but not prolactin) shows strong non-specific adsorption to the immunoadsorbent even in the absence of anti-BiP antibodies. Incomplete preprolactin chains, whether translocated or not, cannot be detected in the immunoprecipitate, with or without anti-BiP antibodies.

In interpreting the above results, it is important to acknowledge that the technique of immune precipitation used would only detect BiP complexes of high affinity and that BiP might

also enter into rapidly dissociable interactions with proteins that we would not have detected. The conclusion that BiP specifically binds to aberrant proteins, and not to native proteins or to nascent polypeptides, is therefore only true for the high affinity interactions described in this study. The *in vivo* experiments described earlier^{1–4,7}, have the same limitations.

If BiP played an essential role in folding or disulphide bond formation of normal proteins, it should bind more extensively to nascent than to mature chains. Because a high affinity interaction with BiP was not observed for nascent prolactin chains, we conclude that it cannot be a necessary step in the folding of all proteins as they enter the ER. We have shown that BiP can bind tightly to aberrantly folded or glycosylated proteins *in vitro* so why are nascent prolactin chains not recognized as aberrant? As incompletely synthesized proteins would not have a native conformation, conformation alone cannot be the feature that BiP recognizes on aberrant proteins. It has been suggested that BiP recognizes surface sites of high hydrophobicity^{3,4}. To reconcile such a model with the data presented here, the site which BiP recognizes on misfolded prolactin chains must be inaccessible on nascent chains. The peptide bands seen in Fig. 3, lanes 11–15, which shows nascent chains, are generated by a poorly understood 'pausing' mechanism. At least 70% of the two predominant bands, 87 and 120 amino acids long, are undergoing translocation (Siegel and P.W., manuscript in preparation) so that 50 amino acids which span the ribosome and the ER membrane¹⁸ would be inaccessible for BiP binding. The N-terminal 70 amino acids that are accessible, however, clearly do not associate with BiP. If BiP were to bind to a site within the last 50 amino acids of prolactin, this site would be unavailable until after synthesis and translocation were completed. Another possibility is that BiP binds only to aggregates of proteins in the ER. This possibility has been previously suggested based on similarity between BiP and other members of the HSP-70 family^{4,19} which are proposed to dissolve protein aggregates (such as those which might form during heat shock). Incorrectly glycosylated or folded proteins might be more likely to aggregate than native molecules²⁰, whereas nascent chains might be prevented from aggregating by their association with the translocation machinery in the ER membrane.

In this *in vitro* system only a fraction of the polypeptide chains that have incorrect disulphide bonds or are not glycosylated are recovered in the BiP immune precipitates. The low recovery could be due to inefficient association with BiP in our *in vitro* translation-translocation system, or dissociation of BiP complexes during the immune precipitation procedures. Alternatively, BiP might recognize only a subset of the aberrant molecules. The protein aggregation model^{4,19}, for example, might predict that only aggregates of aberrant proteins would be bound by BiP.

We have used an *in vitro* system to study the interaction of BiP with newly synthesized proteins so that we could manipulate the conditions of protein synthesis, glycosylation and folding. In addition it eliminates other variables such as changing levels of BiP due to induction and the transport of proteins out of the ER. Our data show that BiP can distinguish between native and aberrant proteins and can form tight complexes with misfolded or aberrantly glycosylated polypeptides *in vitro*. What is the function of such associations *in vivo*? Several observations suggest that BiP is not merely an intracellular garbage collector. Bole *et al.*² have presented evidence that immunoglobulin heavy chains and incompletely assembled oligomers associate transiently with BiP during normal immunoglobulin assembly. It has also been reported that isolated BiP-heavy-chain complexes can be disrupted *in vitro* with ATP⁴. Consistent with this observation, we found it necessary to deplete the ATP in the translation reaction before microsome lysis in order to isolate BiP-protein complexes from our *in vitro* system; if microsomes were lysed without depleting ATP, no peptides co-precipitated with BiP (not shown). It is hard to reconcile an energy-dependent

dissociation of BiP with a model in which BiP's sole function is to bind to aberrant proteins. BiP's characteristic association with aberrant proteins is more likely to be due to a futile attempt to correct their aberration than a preference for their company. The availability of an *in vitro* system should help to elucidate the role of this highly conserved protein.

We thank Dr David Bole for providing monoclonal antibody against BiP, and members of the Kelly lab for critical reading of the manuscript. This work was supported by NIH grants to P.W. and R.B.K.

Received 25 January; accepted 30 March 1988.

- Haas, I. G. & Wabl, M. *Nature* **306**, 387-389 (1983).
- Bole, D. G., Hendershot, L. M. & Kearney, J. F. *J. Cell Biol.* **102**, 1558-1566 (1986).
- Gething, M.-J., McCammon, K. & Sambrook, J. *Cell* **46**, 939-950 (1986).
- Munro, S. & Pelham, H. R. B. *Cell* **46**, 291-300 (1986).
- Pouyssegur, J., Shiu, R. P. C. & Pastan, I. *Cell* **11**, 941-947 (1977).
- Olden, O., Pratt, R. M., Jaworski, C. & Yamada, K. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 791-795 (1979).
- Dorner, A. J., Bole, D. G. & Kaufman, R. J. *J. Cell Biol.* **105**, 2665-2674 (1987).
- Sciandra, J. J., Subjeck, J. R. & Hughes, C. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4843-4847 (1984).
- Kelley, P. M. & Schlesinger, M. J. *Cell* **15**, 1277-1286 (1978).
- Whelan, S. A. & Hightower, L. E. *J. Cell Physiol.* **125**, 251-258 (1985).
- Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 852-862 (1975).
- Lau, J. Y. T. et al. *J. Biol. Chem.* **258**, 15255-15260 (1983).
- Sheele, G. & Jacoby, R. J. *J. Biol. Chem.* **257**, 12277-12282 (1982).
- Kaderbhai, M. A. & Austen, B. M. *Eur. J. Biochem.* **153**, 167-178 (1985).
- Bergman, L. W. & Kuehl, W. M. *J. Biol. Chem.* **254**, 8869-8876 (1979).
- Bergman, L. W. & Kuehl, W. M. *J. Biol. Chem.* **254**, 5690-5694 (1979).
- Hobden, A. H. & Cundliffe, E. *Biochem. J.* **170**, 57-61 (1978).
- Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835-851 (1975).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Gibson, R., Schlesinger, S. & Kornfeld, S. *J. Biol. Chem.* **254**, 3600-3607 (1979).
- Hansen, W., Garcia, P. D. & Walter, P. *Cell* **45**, 397-406 (1986).
- Erickson, A. H. & Blobel, G. *Meth. Enzym.* **96**, 38-50 (1983).
- Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 84-93 (1983).
- Gilmore, R. & Blobel, G. *Cell* **42**, 497-505 (1985).

Intraorganellar calcium and pH control proinsulin cleavage in the pancreatic β cell via two distinct site-specific endopeptidases

Howard W. Davidson*, Christopher J. Rhodes & John C. Hutton†

Department of Clinical Biochemistry,
University of Cambridge, Cambridge CB2 2QR, UK

Insulin is produced from an inactive precursor, proinsulin, through initial endoproteolytic cleavage at sites marked by pairs of basic amino-acid residues^{1,2}. We report here that lysates of insulin secretory granules contain two distinct Ca-dependent acidic endoproteases; one (type I) cleaving exclusively on the C-terminal side of Arg 31.Arg 32 (B-chain/C-peptide junction), the other (type II) preferentially on the C-terminal side of Lys 64.Arg 65 of proinsulin (C-peptide/A-chain junction). The Ca and pH requirements of these proteinases suggested that the type-II proteinase would be active in the Golgi apparatus and the secretory granule, whereas type-I activity would be compatible only with the intragranular environment. Kinetic analyses of (pro)insulin conversion intermediates in [³⁵S]methionine-pulsed rat islets support this supposition. Our results suggest a simple mechanism whereby different dibasic sites can be cleaved in different cellular compartments. In conjunction with the regulation of the ionic composition of such compartments and the operation of post-Golgi segregation, our results also suggest how proteolytic conversion of diverse proproteins destined for different cellular sites can occur differentially and in a regulated manner.

Previous studies have shown that the processing of [¹²⁵I]-labelled human proinsulin³ and of biosynthetically labelled chromogranin A (ref. 4) by lysed subcellular fractions of a transplantable insulinoma is initiated by a novel Ca²⁺-dependent endopeptidase activity which is localized principally to the secretory granule compartment. The activity in purified insulinoma granules was solubilized by sonication in media containing 0.5% Triton X-100 and subjected to chromatography on DEAE-cellulose. Fractions were collected directly onto 'spun' columns of G-25 Sephadex equilibrated with 50 mM acetate buffer pH 5.5 to avoid inactivation of these enzymes by anions contained in the eluting buffer⁴. Two distinct endopeptidase activities were observed (Fig. 1). The first (type-I activity) which eluted with the column void converted [¹²⁵I]proinsulin to 32,33-split proinsulin (Fig. 2). The second (type-II activity) which eluted between 0.25-0.3 M NaCl converted the substrate principally to 65,66-split proinsulin and, to lesser extent, 32,33-split proinsulin and diarginylinsulin (insulin with the B-chain extended by Arg 31.Arg 32). The relative ratio of the latter three products produced by different fractions over the peak of type II activity (Fig. 1) remained essentially constant, a finding consistent with the presence of a single enzyme which cleaves both dibasic sequences in proinsulin but the A-chain/C-peptide junction preferentially.

Identification of the split proinsulin intermediates was on the basis of their co-elution on high pressure liquid chromatography (HPLC) with standards, their conversion by carboxypeptidase B or H via C-terminally extended monobasic forms to des 31,32- and des 64,65-proinsulin respectively, and their conversion to insulin and diarginyl insulin respectively by tryptic proteolysis (30 min at 30 °C with 10 ng trypsin in 100 μ l 50 mM Tris formate pH 7.5; results not shown). These assignments were confirmed by alkaline urea gel electrophoresis and acid urea-gel electrophoresis as appropriate. The two endopeptidase activities when recombined, generated diarginylinsulin in addition to the split proinsulins as expected. This was a similar spectrum of intermediates as generated by limited trypsinolysis of [¹²⁵I]proinsulin and also by lysed secretory granule preparations in which the endogenous carboxypeptidase H was inhibited by adding 1 μ M guanidinoethanemercaptosuccinic acid³. The combination of the two endopeptidase activities with carboxypeptidase H was required to produce insulin. The yield of insulin under suitable conditions reached 80% and there was no evidence of further degradation of the insulin molecule upon extended incubation (16-24 h).

The type I and II endopeptidase activities prepared as in Fig. 1 but in the absence of protease inhibitors were shown to be insensitive to group-specific inhibitors of serine (1 mM diisopropylfluorophosphate), thiol (1 mM iodoacetamide) and aspartyl proteinases (1 mM pepstatin A). Both were inhibited by the chelating agents EDTA and 1,2-cyclohexanediaminetetraacetic acid (CDTA) but not by the heavy metal chelator 1,10-phenanthroline. Following removal of the chelator by gel filtration, activity was completely restored by addition of 5 mM Ca²⁺. But whereas the Lys 64.Arg 65-directed type II activity was reactivated by micromolar Ca²⁺ concentrations ($K_{0.5}$ = 100 μ M), the Arg 31.Arg 32-directed type I activity required millimolar Ca²⁺ ($K_{0.5}$ = 2.5 mM) (Fig. 3). The pH optimum of both activities was 5.5; however, the type II activity was distributed over a broader pH range (Fig. 3). This meant that at pH 7.5 type II activity was around 30-40% of that at pH 5.5, whereas the type I activity was negligible. These data suggested that both activities would be maximal in the ionic environment of the secretory granule where the pH approximates 5.5 (ref. 5) and the ionized Ca²⁺ concentration is 1-10 mM (ref. 6). In the Golgi apparatus and the condensing vacuole, through which these enzymes presumably pass en route for the granule, only the type II activity would be significant as the pH here is closer to neutrality⁷ and Ca²⁺ is present at a lower concentration⁸.

These findings raised the question of whether cleavage of the

* Present address: Department of Biochemistry, Medical Sciences Institute, Dundee, DD1 4HN, UK

† To whom correspondence should be addressed.

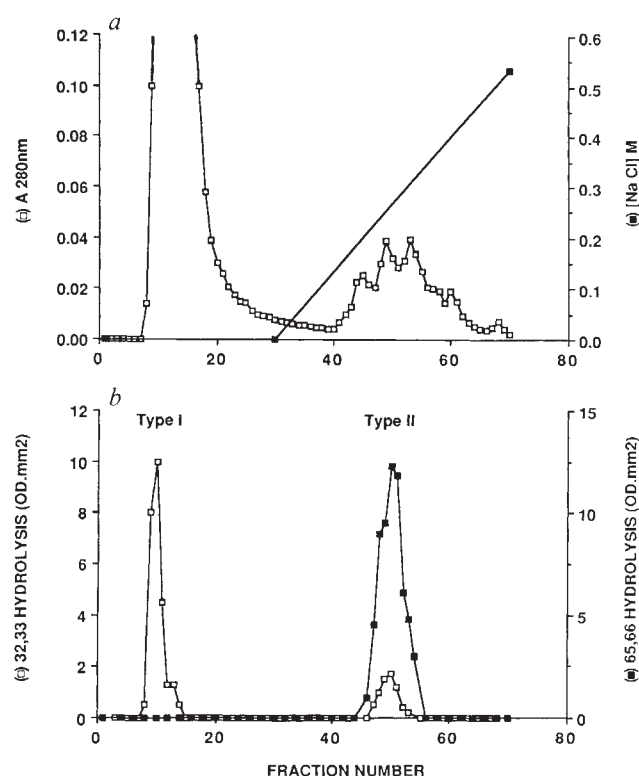


Fig. 1 Separation of insulin secretory granule type-I and type-II endopeptidase activity by ion exchange chromatography. *a*, The $A_{280\text{nm}}$ and the NaCl concentration gradient is shown together with *b*, the elution profiles of endopeptidase activities directed at Arg 31-Arg 32 and Lys 64-Arg 65 sequences of proinsulin.

Methods. All steps were performed at 4°C unless specified. Transplantable rat insulinoma tissue (3–5 g) was homogenized in 270 mM sucrose, 10 mM morpholinoethanesulphonic acid (MES)-KOH, 1 mM EGTA solution, pH 6.5, centrifuged for 6 min at 1,770g and the supernatant (28 ml) layered on to a two-step density gradient composed of 5 ml each of 10% and 20% Nycodenz in 10 mM MES-KOH, pH 6.5. After centrifugation for 2 h at 100,000g, secretory granules were recovered from the 10/20% Nycodenz interface and samples (10–20 mg protein) sonicated for 15 s in 4–5 ml of 20 mM bis-Tris, 0.5% Triton X-100, 1 mM EDTA, 0.1 mM tosyl-phenylethyl-chloromethyl-ketone (TPCK), 50 μM E459, 10 μM Pepstatin A, pH 5.5. The supernatant obtained after centrifugation for 15 min at 35,000g was applied at 0.5 ml min⁻¹ to a 8×40 mm column of DE-52 cellulose equilibrated in 20 mM bis-Tris pH 5.5, the column washed and then eluted with a 0–0.6 M NaCl gradient in the same buffer. The buffer in each fraction (1 ml) was immediately exchanged with 50 mM Na acetate buffer pH 5.5 using a 'spun' column procedure with G25 Sephadex. The assay was performed as follows. Samples (20 μl) were incubated for 2 h at 30°C in 100 μl (final volume) 50 mM Na acetate, 5 mM CaCl₂, 10 μM E-64, 10 μM pepstatin A, 0.1 mM TPCK, pH 5.5 containing 30,000 disintegrations per minute (d.p.m.) [¹²⁵I]proinsulin (15–20 ng). The reaction was terminated by the addition of 12.5 μl 1 M Tris-Cl pH 8, then a 15 min incubation performed at 30°C with diisopropyl-fluorophosphate (DFP)-treated carboxypeptidase B (125 ng in 12.5 μl 50 mM Tris-Cl pH 7.5) to remove newly generated C-terminal basic amino-acid residues. The reaction products were analysed by alkaline-gel electrophoresis and autoradiography as previously described⁵. Activity was quantified by densitometric scanning of bands which were identified as des 31,32 proinsulin (Arg 31-Arg 32-directed), des 64,65 proinsulin (Lys 64-Arg 65-directed activity) and insulin (combined activity).

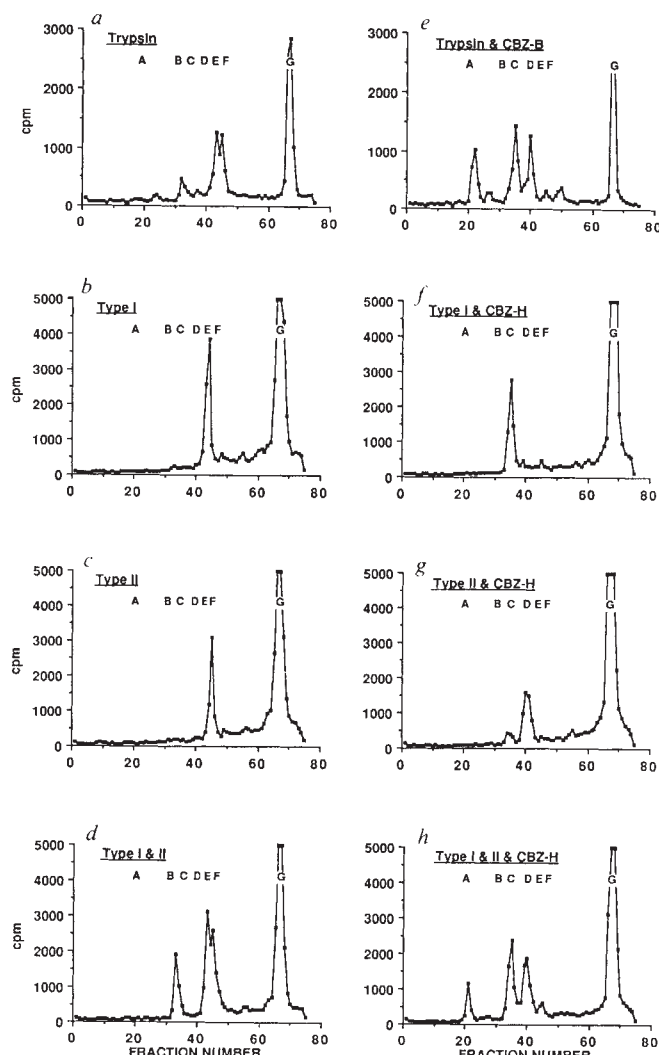


Fig. 2 HPLC analysis of reaction products from proinsulin partially digested with the granule type I and II endopeptidases or with trypsin showing the effects of carboxypeptidase B or H addition. The elution positions of standard human insulin (A), diarginylinsulin (B), des 31,32 proinsulin (C), des 64,65 proinsulin (D), split 32,33 proinsulin (E), split 65,66 proinsulin (F) and proinsulin (G) are indicated. The endopeptidase activity eluting in the DE52 column void (type I activity) and the activity eluting between 0.25–0.3 M NaCl (type II activity) were incubated under the standard assay condition (legend Fig. 1) with the inclusion of either 100 ng purified carboxypeptidase H or of 5 μM guanidinoethanemercaptosuccinic acid to inhibit residual endogenous carboxypeptidase H activity. Digestion of proinsulin (200 μg ; 30,000 d.p.m.) with 20 ng TPCK-treated trypsin was performed in 200 μl 50 mM Tris formate pH 7.5 for 30 min at 30°C, followed by heat inactivation (5 min at 75°C) and a further incubation for 30 min with 2.5 μg carboxypeptidase B as indicated. The HPLC analysis was carried out as follows. Samples were run at 1 ml min⁻¹ on a 250×4.6 mm, 5 μm Apex ODS (octadecylsilica) column which was equilibrated with 50 mM H₃PO₄, 100 mM NaClO₄, 10 mM heptanesulphonic acid, pH 3, containing 33% (v/v) acetonitrile. The column was developed isocratically for 22.5 min then with a linear gradient to 38.5% acetonitrile over 54 min. The gradient fractions (0.6 ml) were collected for gamma counting. Standards were obtained from Eli Lilly or prepared as previously described⁹.

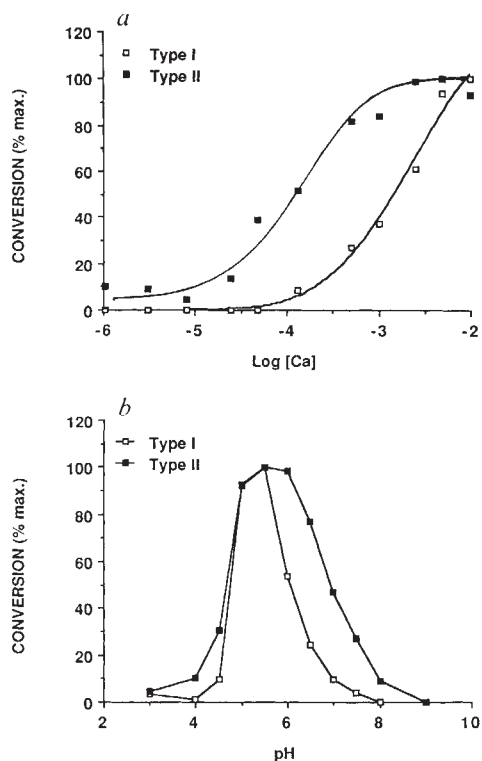


Fig. 3 Ca^{2+} and pH sensitivity of the type I and II granule endopeptidase activities. Assays were performed essentially as in Fig. 1 except that in Ca^{2+} studies the assay contained 2 mM EDTA and variable concentrations of CaCl_2 and for pH studies a buffer system composed of 30 mM Na acetate, 30 mM MES and 30 mM *N*-ethylmorpholine adjusted with HCl was used. Ionized Ca was computed from reference stability constants using an iterative calculation method²⁷.

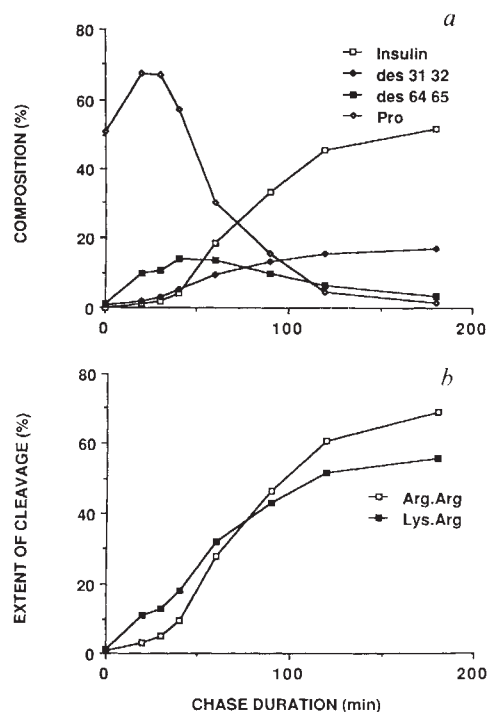


Fig. 4 Proinsulin processing in intact rat pancreatic islets. Islets prepared by collagenase digestion²⁸ were pre-incubated at 37 °C in groups of 100 in 1 ml Krebs bicarbonate medium containing 16.7 mM glucose, then in 100 μl of the same media with 100 μCi [^{35}S]methionine (Amersham; 800 Ci mmol⁻¹) for 5 min, followed by a chase incubation for the times indicated in 1 ml media without radioisotope but containing 2.8 mM glucose and 1 mM methionine. Insulin-related peptides were immunoprecipitated using a solid phase monoclonal anti-human proinsulin antibody (3B7) (ref. 19) and subjected to HPLC as in Fig. 2 but using 34% acetonitrile in the isocratic segment followed by a 34–39.5% acetonitrile gradient. *a*, Radioactivity in each fraction of the gradient (0.6 ml) was determined by liquid scintillation counting, the identifiable peaks of radioactivity integrated and then expressed as a percentage of the immunoprecipitable radioactivity incorporated at the 20 min chase point. (30,000–50,000 d.p.m.). *b*, The extent of cleavage at each site of proinsulin was derived from Fig. 4*a* by summation of the incorporation of radioactivity into insulin and the respective des dibasic proinsulin intermediates. Qualitatively identical results to those shown were obtained on three separate occasions.

dibasic sites in proinsulin in the intact cells occurs sequentially in different compartments and hence whether cleavage at the C-peptide/A-chain junction might be signal which directs segregation of the prohormone to the secretory granule. To examine this, isolated rat pancreatic islets were labelled for 5 min with [^{35}S]methionine, subjected to chase incubations of varying duration and the insulin-related peptides immunoprecipitated, separated by HPLC and their radioactivity determined. In agreement with previous studies¹ insulin was formed 30–40 min post labelling and conversion proceeded to virtual completion over the ensuing 120 min (Fig. 4*a*). The intermediate des 31,32 proinsulin appeared in the islets at the same time as insulin suggesting that it was formed in the same compartment as insulin. Des 64,65 proinsulin, by contrast, appeared much earlier. Intermediates corresponding to the split proinsulins or diarginylinsulin were not observed at any time point; however, these would be expected to have a half life of a few seconds given the high level of carboxypeptidase H in the tissue⁹.

The extent of cleavage of the C-peptide/A-chain junction and B-chain/C-peptide junction during the chase incubation was calculated by the summation of the islet content of the respective radiolabelled des-proinsulin intermediate with radiolabelled insulin (Fig. 4*b*). This analysis indicated that Lys 64.Arg 65-directed cleavage proceeded for 120 min at a linear conversion rate equivalent to 0.43% of the initial radioactivity per min and was initiated soon after the translocation of proinsulin into the endoplasmic reticulum/Golgi (x -axis intercept at $t = -1.5$ min). Arg 31.Arg 32-directed activity also initially proceeded in a linear fashion but with a delay of 21–26 min (by linear extrapolation) and a somewhat higher rate (0.64% min⁻¹). The relative abundance of the des 31,32 proinsulin to des 64,65 proinsulin

at 180 min was similar to the 3:1 ratio reported in normal human pancreas¹⁰.

The delay of ~30 min in Arg 31.Arg 32-directed cleavage of proinsulin and insulin formation coincides with the transfer of labelled proteins to the secretory granule compartment as demonstrated by autoradiography of pulse-chase labelled rat pancreatic islets analysed at the ultrastructural level¹¹ and immunoelectronmicroscopic studies performed with a monoclonal antibody recognizing an epitope which spans the B-chain/C-peptide junction⁷. We propose that a critical event in the initiation of processing at this B-chain/C-peptide junction occurs as a consequence of the insertion or activation of the insulin granule proton translocating ATPase^{12,13} and Ca^{2+} transport proteins¹⁴ in the secretory granule membrane. Granules in which processing has occurred can be distinguished from their immediate clathrin-coated precursors by their more acidic interior^{7,11}. Such granules also appear to be the major morphological sites of Ca^{2+} accumulation^{8,15}.

The stringent requirement for high Ca^{2+} and an acidic pH for the Arg 31.Arg 32-directed processing activity may be a mechanism by which insulin production is limited to the compartment in which it is stored, an important consideration given

that the Zn-insulin complex is less soluble than that of the Zn-proinsulin complex¹⁶. Because Ca^{2+} accumulation in insulin secretory granules in the intact cell is stimulated by insulinotropic stimuli^{8,15,17} there is the further possibility that Ca^{2+} might regulate intragranular proteolysis via this catalytic activity. This could apply in particular to some of the other secretory granule proteins which are synthesized as higher molecular weight precursors^{18,19} which might have a function either within the granule or extracellularly as regulatory factors. This hypothesis is supported by the observation that calcium depletion of islets in culture results in a defect in maturation of newly formed granules consistent with a failure in either insulin formation or its crystallization²⁰. Also, glucose, which stimulates granule Ca^{2+} accumulation¹⁷, is reported to accelerate proinsulin conversion²¹.

The appearance of des 64,65 proinsulin at a time when the bulk of newly synthesized secretory proteins are within the Golgi apparatus of the pancreatic β cell¹¹ is consistent with the postulate that the type II endopeptidase activity can be expressed in a cellular compartment with a neutral pH and a lower Ca^{2+} concentration than the secretory granule. On the basis of the relative rates of cleavage of the Lys 64.Arg 65 and the Arg 31.Arg 32 sites by type II activity (Fig. 2) and the observed conversion of 12% of labelled proinsulin to des 64,65 proinsulin after a 20-30 min chase incubation in islets (Fig. 4), it is estimated that the expression of type II activity in the Golgi would convert only ~2.5% of proinsulin to insulin. Generation of insulin at this intracellular site conceivably accounts for the low rate of constitutive hormone release in islets²².

The processing of proinsulin at the A-chain/C-peptide junction in the Golgi does not appear to be a prerequisite for segregation of the hormone, as Lys 64.Arg 65-directed cleavage continued long after labelled proteins have passed from the Golgi to the secretory granule. The observed generation of des 31,32 proinsulin at the level of the secretory granule further supports this conclusion as this intermediate could only have been produced from proinsulin. By the same argument, processing directed at Lys 64.Arg 65 site is clearly not a prerequisite for processing at the Arg 31.Arg 32 site.

In agreement with published morphological data²³ we conclude that the expression of type II activity in the Golgi, in quantitative terms, is not important for insulin production. However, the activation of the enzyme at this level may be essential to the processing of the precursor forms of proteins which are secreted constitutively, segregated to organelles other than the secretory granule or targeted to the plasma membrane. The relative preference for the Lys.Arg relative to the Arg.Arg sequence of type-II activity reflects the overall frequency of sites of cleavage in different proprotein precursors¹. That the type-II enzyme has a broad protein substrate specificity is suggested by the similarity between this activity and that involved in chromogranin A processing in the pancreatic β cell⁴. Such activity is unlikely to be restricted to pancreatic β cells as proinsulin can be processed correctly by heterologous endocrine cells²⁴ and in distantly related eukaryote organisms²⁵. Likewise, endopeptidase activity with properties similar to those observed here is involved in proalbumin processing by hepatocyte microsomes²⁶.

It is proposed that two distinct proprotein proteolytic activities exist whose substrate specificity is defined by the primary sequence of the dibasic site. Their activities appear to be regulated by compartmental pH and Ca^{2+} , which provides a simple mechanism whereby different dibasic sites within the one protein can be cleaved in different subcellular compartments. Additional factors regulating cleavage would be the segregation or partitioning of the proprotein substrates and the processing enzymes at the level of the trans Golgi and the control of ionic gradients established in the various compartments representing the pathway taken by the precursor molecules. The interaction of these elements provides a versatile mechanism whereby processing of

diverse proproteins destined for different cellular sites can occur both differentially and in a regulated manner.

We thank Mrs M. Peshavaria for technical assistance, Dr B. H. Frank (Lilly) for [¹²⁵I]proinsulin and standards and Dr K. Hanada (Taisho) for E-459. This work was supported by the Wellcome Trust, British Diabetic Association and the MRC. C.J.R. is a Juvenile Diabetes Foundation Research Fellow.

Received 8 February; accepted 28 March 1988.

1. Docherty, K. D. & Steiner, D. F. *A. Rev. Physiol.* **44**, 625-638 (1982).
2. Loh, Y. P., Brownstein, M. J. & Gainer, H. A. *Rev. Neurosci.* **7**, 189-222 (1984).
3. Davidson, H. W., Peshavaria, M. & Hutton, J. C. *Biochem. J.* **246**, 279-286 (1987).
4. Hutton, J. C., Davidson, H. W. & Peshavaria, M. *Biochem. J.* **244**, 457-464 (1987).
5. Hutton, J. C. *Biochem. J.* **204**, 171-178 (1982).
6. Hutton, J. C., Penn, E. J. & Peshavaria, M. *Biochem. J.* **210**, 297-305 (1983).
7. Orci, L. *et al. Cell* **49**, 865-868 (1987).
8. Herman, L., Sato, T. & Hales, C. N. *J. Ultrastruct. Res.* **42**, 298-311 (1973).
9. Davidson, H. W. & Hutton, J. C. *Biochem. J.* **245**, 575-582 (1987).
10. Given, B. D. *et al. J. clin. Invest.* **76**, 1398-1405 (1985).
11. Orci, L. *Diabetologia* **28**, 528-546 (1985).
12. Hutton, J. C. & Peshavaria, M. *Biochem. J.* **204**, 161-170 (1982).
13. Rhodes, C. J., Lucas, C. A., Mutkoski, R. L., Orci, L. & Halban, P. A. *J. biol. Chem.* **262**, 10712-10717 (1987).
14. Formby, B., Capito, K., Egeberg, J. & Hedekov, C. J. *Am. J. Physiol.* **230**, 441-448 (1976).
15. Howell, S. L. & Tyhurst, M. *J. Cell Sci.* **21**, 415-422 (1976).
16. Grant, P. T., Coombs, T. L. & Frank, B. H. *Biochem. J.* **126**, 433-440 (1972).
17. Rhodes, K.-D., Hahn, H.-J., Gylfe, E., Borg, H. & Hellman, B. *Molec. cell. Endocrinol.* **16**, 205-220 (1979).
18. Hutton, J. C., Davidson, H. W., Grimaldi, K. A. & Peshavaria, M. *Biochem. J.* **244**, 449-456 (1987).
19. Grimaldi, K. A., Siddle, K. & Hutton, J. C. *Biochem. J.* **245**, 567-573 (1987).
20. Howell, S. L., Tyhurst, M., Duvefelt, H., Andersson, A. & Hellerström, C. *Cell Tissue Res.* **188**, 107-118 (1978).
21. Nagamatsu, S., Bolaffi, J. L. & Grodsky, G. M. *Endocrinology* **120**, 1225-1231 (1987).
22. Rhodes, C. J. & Halban, P. A. *J. Cell Biol.* **105**, 145-153 (1987).
23. Steiner, D. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6184-6188 (1987).
24. Moore, H.-P., Walker, M. D., Lee, F. & Kelly, R. B. *Cell* **35**, 531-538 (1983).
25. Thim, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 6766-6770 (1986).
26. Brennan, S. O. & Peach, R. J. *FEBS Lett.* **229**, 167-190 (1988).
27. Storer, A. C. & Cornish-Bowden, A. *Biochem. J.* **159**, 1-5 (1976).
28. Lacy, P. E. & Kostianovsky, M. *Diabetes* **16**, 35-39 (1967).

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should bear in mind that the journal is read by scientists for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50-80 words written to advertise their contents in general terms, which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, many are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inadequate funds to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unreferenced abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963-65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not required.

Velocity sedimentation of cells

T.G. Pretlow and T.P. Pretlow

Velocity sedimentation provides a variety of methods for the separation of cells according to their rates of sedimentation. Many approaches are available.

SEDIMENTATION in a continuous gradient is an ideal analytical or preparative method for the separation of molecules and particles according to their physical properties. The two most widely used forms of sedimentation are isopycnic and velocity sedimentation. Isopycnic, or density equilibrium, sedimentation is carried out with a sufficient product of force and time to permit the material being centrifuged to arrive at its intrinsic density in the gradient. Velocity sedimentation separates particles by their rates of sedimentation and is used for the characterization of macromolecules¹.

We shall address the velocity sedimentation of "cells", recognizing that what is said is also applicable to organelles, viruses, fragments of ore, and other particles. The description of sedimentation for cells is simpler than that for molecules in that shape, degree of hydration, diffusion, and charge become negligible for these larger particles². The equation for sedimentation is

$$v = drdt^{-1} = a^2(D_p - D_m)\omega^2r18\eta^{-1}$$

when v = the velocity of sedimentation, r = the distance of the cell from the center of revolution, t = duration of sedimentation, a = diameter of the cell, D_p = density of the cell, D_m = density of the medium or gradient at the location of the cell, ω = angular velocity (speed of centrifugation), and η = viscosity of the medium at the location of the cell. With small modifications for the absence of a density gradient in elutriation or for the use of unit gravity instead of a centrifugal field, this description of sedimentation is applicable to all techniques of velocity sedimentation.

Cells have two physical properties that determine simultaneously their rates of sedimentation: density and diameter. While cells can be separated according to density by isopycnic sedimentation, no method of sedimentation permits the separation of cells according to diameter. The most successful techniques for velocity sedimentation differ from older techniques in that the gradients employed have small slopes ($\text{gml}^{-1}\text{cm}^{-1}$) and are of lower density. The optimal design of such gradients has been discussed³.

There are four commonly employed approaches to the separation of cells by velocity sedimentation: counterstreaming centrifugation or elutriation^{4,5}, sedimentation in a reorienting zonal rotor⁷, sedimentation at unit gravity in either fixed^{8,9} or reorienting^{10,11} chambers, and isokinetic sedimentation¹². The first two techniques

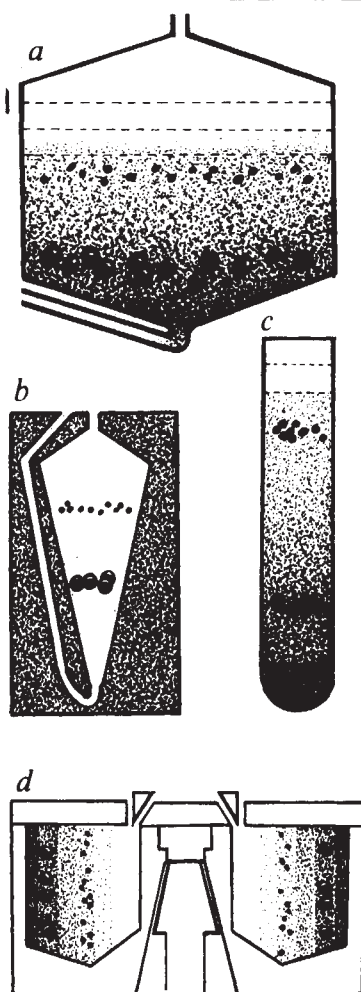


Fig. 1 Devices that are commonly used for velocity sedimentation. Density gradients are represented by variable shading. Rapidly sedimenting and slowly sedimenting cell populations are represented in the devices for comparison. *a*, Device containing a gradient for the separation of cells at unit gravity^{8,9,11}. *b*, Chamber for elutriation^{5,6}. *c*, Tube with 13-cm gradient for isokinetic sedimentation^{2,12}. *d*, Cross-section of a reorienting zonal rotor⁷.

are best suited for the separation of 10^8 – 10^9 cells; the last two techniques, 10^6 – 10^8 cells. All of these techniques can be used under sterile or nearly sterile conditions. Isokinetic sedimentation and sedimentation at unit gravity are the easiest techniques to use under sterile conditions because all of the equipment required for these techniques can be autoclaved. Isokinetic sedimentation requires only equipment that is available in most laboratories, and the equipment required for sedimentation at unit gravity is inexpensive and commercially available. Both elutriation and sedi-

mentation in a reorienting zonal rotor require particular centrifuges with expensive, specific kinds of rotors.

The most common pitfalls encountered in the use of velocity sedimentation result from a lack of familiarity with sedimentation artefacts². In particular, the loss of purity that occurs when the band capacity is exceeded appears not to be widely appreciated. Other drawbacks arise from discontinuous gradients^{2,13} or gradients that are suboptimally designed¹.

Improvements in velocity sedimentation are likely to occur both in the design of apparatus and in the design of specific applications. The resolution that is available with current apparatus is generally more than adequate for particles that are as heterogeneous as mammalian cells. To improve the apparatus, the steeper^{3,14} centrifugal field (force per centimetre) that is obtained as the gradient is moved closer to the centre of the centrifuge may be exploited. Because the wall effect artefact increases toward the centre of the centrifuge, this approach would be more rewarding in zonal rotors and perhaps in elutriators than in parallel-walled tubes. Care must be taken not to position parallel-walled tubes too close to the centre of revolution because the wall effect becomes quite severe. Mixing more than one gradient medium to form a gradient with a viscosity which remains constant or decreases with increasing density could improve the technique. □

Thomas G. Pretlow and Theresa P. Pretlow are at the Institute of Pathology, Case Western Reserve University, Cleveland, Ohio 44106, USA. For more information, fill in reader service number 100.

- Martin, R.G. & Ames, B.N. *J. biol. Chem.* **236**, 1372–1379 (1961).
- Pretlow, T.G., Weir, E.E. & Zettergren, J.G. *Int. Rev. exp. Path.* **14**, 91–204 (1975).
- Pretlow, T.G. *et al. Biochem. J.* **174**, 303–307 (1978).
- Lindahl, P.E. *Nature* **161**, 648–649 (1948).
- Lindahl, P.E. *Biochim. biophys. Acta* **21**, 411–415 (1956).
- Pretlow, T.G. & Pretlow, T.P. *Cell Biophys.* **1**, 195–210 (1979).
- Pretlow, T.P. & Pretlow, T.G. in *Cell Separation: Methods and Selected Applications* Vol. 2 (eds Pretlow, T.G. & Pretlow, T.P.) 221–233 (Academic, New York, 1983).
- Peterson, E.A. & Evans, W.H. *Nature* **214**, 824–825 (1967).
- Hymen, W.C. & Hatfield, J.M. in *Cell Separation: Methods and Selected Applications* Vol. 3 (eds Pretlow, T.G. & Pretlow, T.P.) 163–194 (Academic, New York, 1984).
- Bont, W.S. & Hilgers, J.H.M. *Prep. Biochem.* **7**, 45–60 (1977).
- Wells, J.R. in *Cell Separation: Methods and Selected Applications* Vol. 1 (eds Pretlow, T.G. & Pretlow, T.P.) 169–189 (Academic, New York, 1982).
- Pretlow, T.G. & Pretlow, T.P. in *Cell Separation: Methods and Selected Applications* Vol. 5 (eds Pretlow, T.G. & Pretlow, T.P.) 281–309 (Academic, New York, 1987).
- de Duve, C. *J. Cell Biol.* **50**, 20d–55d (1971).
- Pretlow, T.G. & Boone, C.W. *Science* **161**, 911–913 (1968).

Microbiologists meet in Miami Beach

Freeze dryers, fermenters and special media preparations are only a few of the products to be displayed at the American Society for Microbiology annual meeting next week in Miami Beach, Florida. Over 300 exhibitors will be in attendance.

THE American Type Culture Collection will have information on its **laboratory workshops** for the second half of 1988 available in booth 947 (*Reader Service No. 101*). Topics for the rest of the year include hands-on courses on recombinant DNA technology, microbial fermentation, freezing and quality control of cell cultures and hybridomas, diagnostic virology, and clinical diagnostic gene-probes. There will also be a workshop devoted to the use of computers in microbiology. The classes last from between three and five days, and take place at ATCC's Rockville facility.

The newest version of the **Softmax microplate reader software** has just been released from Molecular Devices Corporation, for use with the Vmax kinetic microplate reader and an IBM PC (*Reader Service No. 102*). The software is designed for the collection, analysis, management and reporting of data. The new negative kinetics feature measures and plots reactions in which optical densities decrease relative to the first data point, and kinetic run times can now be specified in minutes and seconds rather than decimal minutes. Standard curve interpolation has been added to the software, enabling the user to request interpolated X and Y values directly from the standard curve display. Softmax will also now automatically mask overranged raw data values of 4.000 optical density or greater and exclude them from display and analysis. Molecular Devices will be in booths 1223-1224.

Zymed Laboratories will be exhibiting its **monoclonal antibodies to human IgG subclasses** in booth 1546 (*Reader Service No. 103*). Zymed is offering mouse monoclonal antibodies specific for human IgG₁, IgG₂, IgG₃, IgG₄, kappa light chain, lambda light chain, IgA₁, IgA₂, IgM, IgD and IgE. The first five antibodies are also available as conjugates to peroxidase, alkaline phosphatase and biotin. Prices for the unconjugated antibodies average \$50 (US) for 0.2 mg per 0.2 ml.

Spinning, saving and moving

Du Pont will be showing off its newest addition to the Sorvall centrifuge line in booths 1100-1103 and 1148-1151 — the model RC-28S Supraspeed **refrigerated centrifuge** (*Reader Service No. 104*). The \$18,700 (US) Supraspeed centrifuge has a range of speeds that covers the separation spectrum from superspeed to low-end



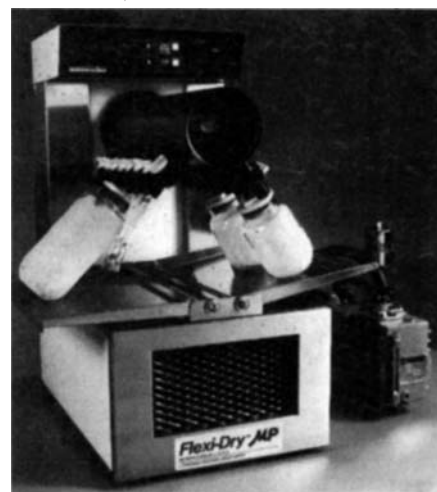
From superspeed to low-end ultraspeed spins. ultraspeed. With a top speed of 28,000 r.p.m. and a top relative centrifugal force of $100,000 \times g$, the Supraspeed eliminates the need to use both a superspeed and an ultraspeed centrifuge for multi-step fractionations. The Supraspeed's Spin-Right rotor management system automatically identifies the rotor, relieving the operator of remembering and entering codes. The Rotor-Scan control system controls temperature, speed and acceleration/deceleration conditions, and uses ultrasonic imaging technology to detect improper installation of the rotor. All existing Sorvall superspeed rotors are compatible with the SupraSpeed; fixed angle and swinging bucket rotors specially developed for use with the SupraSpeed are also available.

Offensive odours from autoclaves can be eliminated with the Odo-clave **auto-clave deodorant discs** from Bel-Art Products (*Reader Service No. 105*). Available in clove, cherry and pine scents, the Odo-clave deodorant discs are packed in both single- and triple-strength formulations. The discs open automatically during the heating cycle, releasing a deodorant which neutralizes the offensive odours and leaves behind a mild residual scent. The disc package ensures no spills, ease of use and easy storage. Bel-Art, exhibiting in booth 203, sells the Odo-clave discs for \$36-56 (US) for a package of 100.

Bio-Rad will be displaying a wide range of products in booths 428-431. One of the

newest items from Bio-Rad is the C/P Lift **colony screening membrane** for the detection of bacterial colonies and plaques (*Reader Service No. 106*). Bio-Rad says the membrane has a higher binding capacity and higher tensile strength than conventional nitrocellulose membranes, and offers a high signal to noise ratio with sharp, distinct colony signals. The nylon-based membrane reduces the denaturing and fixing of nucleic acids to the membrane to a two-step process, eliminating baking and prehybridization steps. Positively hybridized colonies appear as sharp and well-defined signals with low background signal, says Bio-Rad.

In booths 835-836, FTS Systems, Inc. will be introducing its 3-litre microprocessor controlled **freeze dryer**, the Flexi-Dry (*Reader Service No. 107*). FTS Systems says the freeze dryer is capable of condensing 3 litres of vapour in 24 hours and is available in -55°C or -85°C refrigeration configurations. The microprocessor activates the refrigeration and vacuum systems in the correct sequence and then signals the operator when the system is



The 3-litre Flexi-Dry freeze dryer from FTS Systems can condense 3 litres in 24 hours.

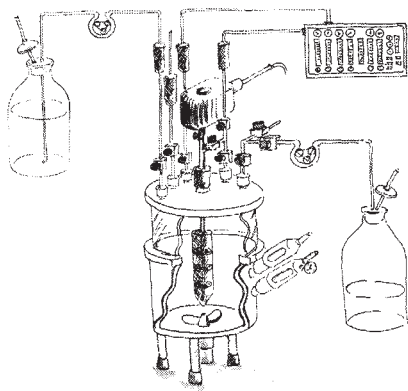
ready to accept the addition of product. If freeze drying conditions are not met at the initial start up or are lost during system operation, the microprocessor indicates which lyophilization parameters have not been obtained. The \$4,900-7,300 (US) Flexi-Dry freeze drying system is also available with an optional titanium condenser; stainless steel, Teflon or Lucite 16-port manifolds; and standard or corrosion-resistant vacuum pumps.

The Rainin Instrument Company recently introduced the model EDP-2 battery-operated **motorized pipette** (Reader Service No. 108). The \$235 (US) EDP-2 has a pipette mode and a dispense mode, and the speed is adjustable from regular to fast. Operating parameters are entered into an internal microcomputer using control keys on a simple keypad, and the current status and volume setting are shown on a liquid crystal display. An electronic linear actuator, controlled by the internal microcomputer, performs all liquid measurement operations. Rainin's pipette is powered by a long-life lithium battery, supplied with the pipette, which Rainin says powers the instrument through over 100,000 pipetting cycles. Six sizes of the model EDP-2 pipettes are available, with volumes ranging from 10 to 2,500 μ l. Rainin Instrument Company will be exhibiting in booths 619-621, and 630-632.

To grow in...

Sarstedt, in booths 1709-1710, will be showing its new line of **shakers and mixers** for tubes, test plates, bottles and dishes (Reader Service No. 109). The \$626 (US) model TPM-2 mixer is designed to provide careful and thorough mixing of samples, and can be adapted to different test requirements using its variable speed and timer controls. The mixer's loading plate has a non-skid surface that Sarstedt says keeps test plates and dishes firmly in place. Its surface area can accommodate a variety of test plates or a 9 x 11 inch dish for destaining gels. The \$575 (US) model CM-9 shaker/mixer has nine different interchangeable trays to hold microtubes, petri dishes, bottles, beakers, Erlenmeyer flasks, round-bottomed flasks, tissue culture containers or test tubes.

Chemap, in booths 436-437, has introduced a new, patented method for the **cultivation of animal cells** which combines



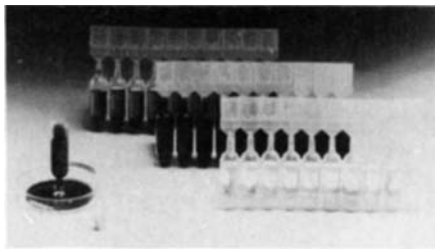
Chemap's ChemCell bioreactor system.

the advantages of bubble-free aeration with cell-free harvesting of media — the ChemCell system (Reader Service No. 110). In Chemap's ChemCell system, cells are grown inside a fine-meshed screen bioreactor cylinder that is im-

mersed into the fermenter. The gas mixture is introduced at the bottom of the fermenter, and vibration of the cylinder breaks down the gas bubbles as they rise to the top. Cell-free medium can be harvested directly by tapping the fermenter contents while fresh medium is added, resulting in a continuous perfusion culture system. Chemap says the ChemCell system has shown good results in the growth of both hybridoma cells in suspension and anchorage-dependent cells on microcarriers. Processes can be scaled up simply by increasing the size of the meshed vibrating cylinder, or by using more than one cylinder in the same fermenter vessel.

Media matters

Millipore is now offering a variety of pre-sterilized growth culture **media in plastic ampoules** for culturing microorganisms on



Millipore sells a variety of types of media in its disposable plastic ampoules.

membrane filters for microbiological analysis (Reader Service No. 111). The media is packaged in 2 ml ampoules, and can be used with Millipore's petri dishes and monitors. Special media for yeasts and moulds, total bacteria, total coliform, and an orange serum medium for use in analysing beverages containing fruit or citrus juices are available packaged in the ampoules. Millipore is exhibiting in booths 949-951.

Two pre-mixed, pre-weighed **biological buffers** have been added to Eastman Kodak's ready-to-use buffer line (Reader Service No. 112). Kodak's BufferEZE HEPES and BufferEZE MOPS are zwitterionic buffers prepared according to the procedures of Good. The buffers come in individually wrapped foil packets; prices range from \$12 to \$40 (US). Eastman Kodak is exhibiting in booth 7-A.

Jouan, exhibiting in booths 1448-1449, has upgraded its model SH105E and SH110E **media preparators** with new pourer and stacker modules and a programmable peristaltic pump (Reader Service No. 113). The model DM150 pourer accepts all sizes of petri dishes between 50 and 155 mm, round or square. The dose size and rate can be varied to meet specific requirements. The model SM240 stacker accepts 55, 90, 95 and 100 mm plate sizes. The model PM600 peristaltic pump has an alpha-numeric display and 10 memories. □

Japanese Bio/Technology

The journal *Bio/Technology* and Kyoikusha, the Japanese publisher of the popular science magazine *Newton* announced last week plans to translate parts of *Bio/Technology* in Japanese to be distributed with the English version beginning in July.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices are sometimes nominal and apply only in the country indicated.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.48

Medwire Div.

stocks a wide variety of

TEFLON COATED*

Platinum alloys, stainless steel, silver, multistrand and single strand available for prompt shipment. Write for descriptive brochure and price lists.



Medwire Div.

Affiliate of Sigmund Cohn Corp.

123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300

*Reg. T.M. of E.I. DuPont

Reader Service No.49



SINGER

INSTRUMENTS

**micro injection?
micro dissection?
micro surgery?
micro positioning?**



Micromanipulators

Singer Instrument Co. Ltd.

Treborough Lodge, Roadwater, Watchet, Somerset,
England, TA23 0QL. Tel: Washford (0984) 40226,
Telex: 337492 Corncab G

Reader Service No.50

IT demand continues to boom

Richard Pearson

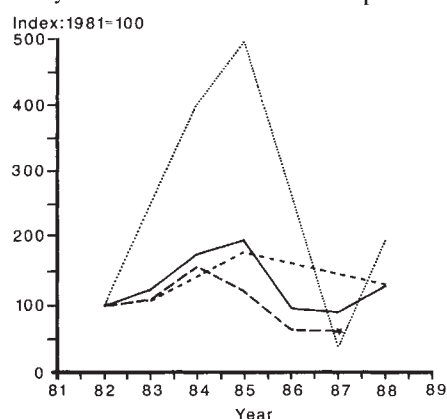
Demand in software market balances cutback in recruitment in the computer and electronics industries.

WITH the UK economic recovery stretching into its sixth year, skill shortages are once again growing. At the beginning of 1988, one in four companies expected shortages of skilled labour to act as a constraint on achieving higher levels of output. This represented a doubling over the previous 12 months and the figure is starting to approach the level recorded in the late 1970s. The occupations affected include the construction trades, machinists, welders and mechanics. The biggest problems, however, relate to professional-level skills and engineers and computing staff in particular.

There are now about 230,000 professional information technology (IT) staff in the United Kingdom, a figure that has been growing at over 5 per cent per annum. While IT skill shortages have persisted in some form for the past decade, their characteristics have changed significantly over the past two years, emphasizing the many discrete occupations associated with IT. Although the demand for different categories of IT specialists is in part driven by technological change, far more important influences have been the take-up of the different aspects of IT by industry and commerce, and the economic growth or otherwise of the users of IT in each sector as well as that of the manufacturers of IT. This is graphically illustrated by the events of the past two years when there were redundancies and massive cutbacks in recruitment in the computer and electronics industries. As these companies are the main recruiters of electronics specialists and of new IT graduates, their cutback had a major impact on these parts of the labour market. For example, IBM, whose graduate intakes had risen in recent years to about 500 in 1985, recruited fewer than 100 the following year. Other major computer and electronics companies that had been forecasting increased needs also cut back, in one case from 140 to 40, in another from 400 to 200. Not all were affected so dramatically but these cutbacks helped to alleviate some of the shortages of new graduates in 1986 and 1987. At the same time, the demand for experienced electronics staff also fell back.

In contrast, the demand by the software houses and the users continued to boom, reflecting the continuing 20 per cent a year growth in the software market. Major growth points included the booming City of London in the run up to deregulation, the finance sector and the rapid take-up of

IT in the retail sector. As a result, the demand for both experienced software staff and graduates to be trainees continued to rise with some software houses continuing to increase their staffing levels by 20 per cent or more, one with under 2,000 employees taking on 250 new graduates as well as more than 300 experienced recruits in a year. Growth in demand was particu-



IT recruitment to major computer and electronics groups, 1981-1988. *, Data for further years not available; dotted line, 1985 intake totalled over 2,500. Source: IT Manpower Monitor 1988, IMS.

larly rapid for networking and communications specialists.

While all sectors looked to graduates for their source of trainees, most recruitment activity centred on experienced staff, two of whom were taken on for each new graduate. Apart from the financial and public staff, there was little retraining or upgrading of existing staff. While the electronics and computer companies were the main recruiters of electronics graduates, computer-science graduates were sought by recruiters from all sectors. But there were still mixed views about the relevance of computer science graduates; some recruiters argued that the courses were too theoretical and so tended to take graduates from any background and to provide training themselves.

Recruitment difficulties continued for those seeking experienced staff. The employers' main response to the problem has been to continue to raise salaries and invest in training, although the amount of training remains low. The earnings of computing staff rose by 12.4 per cent in the year to the end of 1987, well ahead of the national average, and followed rises of 11.9 per cent and 12.5 per cent in the previous two years, again well ahead of

the national averages of around 8 per cent. In October 1987, a trainee programmer was paid £8,693 (up by 18.5 per cent), a senior programmer £15,613 (up by 12.9 per cent) and senior systems analyst £18,538 (up by 10.1 per cent).

Although salaries in higher education have been falling behind the market for some years, recruitment and retention of academic staff has improved in some respects in the past two years. Although 86 per cent of IT departments reported recruitment difficulties in 1987 and the proportion of unfilled vacancies, at just under 10 per cent, remained unchanged, the departments had been able to recruit nearly 300 IT staff, nearly twice as many as in 1985. More encouragingly, wastage rates had fallen to about 7 per cent.

Although demand for new graduates, especially those with electronics skills, had eased in the past two years, the supply has started to improve as a result of the government's IT Initiative and the Engineering and Technology Programme which has been boosting the number of degree places in IT (electronics, computer science) subjects. But there is serious concern that all the extra places from 1988 onwards may not be filled, given the 25 per cent decline in the number of 18-year-olds over the next seven years; some existing courses are already struggling to fill their places, applications to university computer science courses having fallen by 14 per cent over the past three years. Attempts are being made to widen access and boost the demand for places. One of the most significant challenges and opportunities is to increase participation by women, who still account for less than 10 per cent of those on IT courses.

Demand is likely to continue to grow by about 5 per cent a year, provided there is no major economic reverse. There are, however, likely to be significant differences between sectors and consequently for individual IT occupations. Software and networking skills are expected to be the growth points. As the supply of new IT graduates, the prime source of new trainees, will not increase significantly, if at all, skill shortages are likely to continue into the 1990s unless employers improve their use of scarce skills and train and retrain to match the standards of the best employers. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.